

Genomweite Untersuchungen zur Evolution und Methylierung von CpG-Inseln

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
Dr. rer. nat.

der Fakultät für
Biologie
an der

Universität Duisburg-Essen

vorgelegt von
Katrin Grothaus
(geb. Rademacher)

aus Hagen
April 2016

Die der vorliegenden Arbeit zugrunde liegenden Ergebnisse wurden am Institut für Humangenetik der Universität Duisburg-Essen durchgeführt.

1. Gutachter: Prof. Dr. Bernhard Horsthemke
2. Gutachter: Prof. Dr. Daniel Hoffmann
3. Gutachter: —

Vorsitzender des Prüfungsausschusses: Prof. Dr. Ralf Küppers
Tag der mündlichen Prüfung: 28. Juni 2016

Teile der vorliegenden Arbeit wurden bereits in den folgenden Publikationen veröffentlicht:

Grothaus K, Kanber D, Gellhaus A, Mikat B, Kolarova J, Siebert R, Wieczorek D, Horsthemke B. Genome-wide methylation analysis of retrocopy associated CpG islands and their genomic environment. *Epigenetics*. 2016. doi: 10.1080/15592294.2016.1145330

Rademacher K, Schröder C, Kanber D, Klein-Hitpass L, Wallner S, Zeschnigk M, Horsthemke B. Evolutionary origin and methylation status of human intronic CpG islands that are not present in mouse. *Genome Biol. Evol.* 2014, 6(7):1579-88. doi.1093/gbe/evu125

Eine ausführliche Publikationsliste ist im Anhang B.2 angegeben.

Inhaltsverzeichnis

Abkürzungsverzeichnis	viii
1 Einleitung	1
1.1 CpG-Inseln	1
1.2 DNA-Methylierung	2
1.3 Histonmodifikationen	4
1.4 Genomische Prägung	5
1.5 Retrotransposition	9
1.6 Das geprägte <i>RB1</i> -Gen	10
1.7 Zielsetzung	12
2 Material und Methoden	13
2.1 Daten	13
2.1.1 CGI-Daten	14
2.1.2 CCDS-Daten	15
2.1.3 Retrokopie-Daten	15
2.1.4 Gen-Daten (RefSeq)	16
2.1.5 Methylom-Daten	17
2.1.6 Chromatin-Daten	18
2.1.7 Daten über repetitive Sequenzen	18
2.2 <i>In silico</i> Methoden	19
2.2.1 Nummerierung der CGIs	19
2.2.2 Bestimmung der relativen Lage von CGIs zu proteinkodierenden Regionen (CCDS)	20
2.2.3 Sequenzvergleich	23
2.2.4 Evolutionsanalyse	23
2.2.5 Finden von Retrokopie-assoziierten CGIs	24
2.2.6 Analyse der Umgebung von Retrokopien in Hinblick auf repetitive Sequenzen	24
2.2.7 Berechnung der Methylierung von bestimmten Regionen	24
2.2.8 Statistik	25
2.3 Material	25
2.3.1 Chemikalien	25
2.3.2 Standardlösungen	25
2.3.3 Größenstandards	26
2.3.4 Enzyme	26

2.3.5	Oligonukleotide	26
2.3.6	Humane Proben	27
2.3.6.1	Monozyten	27
2.3.6.2	Trios	27
2.3.6.3	Patienten mit einem Multilocus- <i>Imprinting</i> -Defekt (MLID)	27
2.3.6.4	Gewebe	28
2.4	<i>In vitro</i> Methoden	28
2.4.1	<i>Whole-Genome Bisulfite Sequencing</i> (WGBS)	28
2.4.2	Arbeiten mit DNA	28
2.4.2.1	Isolierung von DNA	28
2.4.2.2	Konzentrationsbestimmung von DNA	30
2.4.2.3	Polymerasekettenreaktion (PCR)	30
2.4.2.4	Auftrennung von DNA in Agarosegelen	32
2.4.2.5	Aufreinigung von PCR-Produkten	32
2.4.2.6	Sequenzierung nach Sanger	33
2.4.2.7	Bisulfitkonvertierung von DNA	33
2.4.2.8	Methylierungsanalyse von Bisulfit-behandelter DNA . . .	34
2.4.3	Arbeiten mit RNA	35
2.4.3.1	Isolierung von RNA	35
2.4.3.2	Konzentrationsbestimmung von RNA	36
2.4.3.3	DNase-Behandlung	36
2.4.3.4	Reverse Transkription (RT)	36
2.4.3.5	Amplifikation der cDNA (RT-PCR)	37
2.4.3.6	Auftrennung von RNA in Agarosegelen	37
2.4.3.7	Aufreinigung von PCR-Produkten	37
2.4.3.8	Sequenzierung nach Sanger	37
2.4.3.9	Quantitative Analyse des allelischen Transkriptverhältnisses	38
2.5	Methylierungsanalysen	39
2.5.1	Auswertung Illumina HiSeq2500	39
2.5.2	Auswertung Roche/454 GS Junior	39
3	Ergebnisse	41
3.1	Ist der <i>RB1/Rb1</i> -Locus einzigartig in seinen Eigenschaften?	41
3.1.1	Identifikation von intronischen CGIs im Menschen, die es nicht in der Maus gibt	41
3.1.2	Methylierungsanalysen von human/nicht-murin intronischen CGIs .	43
3.1.2.1	Suche nach allelspezifischer Methylierung	44
3.1.2.2	Methylierungsanalysen mittels Roche/454 GS Junior . . .	47
3.1.3	Evolutionärer Ursprung von human/nicht-murin intronischen CGIs	52
3.2	Genomweite Analysen von Retrokopie-assoziierten CGIs	55
3.2.1	Identifikation von Retrokopie-assoziierten CGIs und deren Klassifi- zierung	55

3.2.2	Methylierungsanalysen von Retrokopie-assoziierten CGIs	57
3.2.2.1	Methylierungsanalysen der Umgebung von Retrokopie-assoziierten CGIs	59
3.2.2.2	Vergleich der Methylierung des ancestralen CGIs mit der Methylierung des R-CGIs	60
3.2.2.3	Suche nach allelspezifischer Methylierung	60
3.2.2.4	Methylierungsanalysen mittels Roche/454 GS Junior . . .	63
3.2.3	Das R-CGI 9261_1_hg19	64
3.2.3.1	Chromatin-Daten zu R-CGI 9261_1_hg19	65
3.2.3.2	Methylierungsanalyse des R-CGIs 9261_1_hg19 in informativen Eltern-Kind-Trios (Blut)	67
3.2.3.3	Methylierungsanalyse des R-CGIs 9261_1_hg19 in Patienten mit einem Multilocus- <i>Imprinting</i> -Defekt (Blut) . . .	68
3.2.3.4	Expressionsanalysen des R-CGIs 9261_1_hg19	70
3.2.3.5	Analyse der benachbarten Gene <i>TRA2A</i> und <i>IGF2BP3</i> .	70
4	Diskussion	73
4.1	Ist der <i>RB1/Rb1</i> -Locus einzigartig in seinen Eigenschaften?	73
4.2	Analyse von Retrokopie-assoziierten CGIs	76
5	Zusammenfassung	81
	Literaturverzeichnis	83
A	Anhang Dissertation	99
A.1	URLs von benutzten Datenbanken und Programmen	99
A.2	Primersequenzen	100
A.2.1	Genomische Primersequenzen	100
A.2.2	Bisulfit Primersequenzen	101
A.2.3	cDNA Primer	101
A.2.4	SNaPshot Primer	102
A.2.5	MID-Primer	102
A.3	Ergebnisse der Genotypisierung	104
A.3.1	Genotypisierung der 22 Monozyten	104
A.4	Ausführliche Ergebnisse der Methylierungsanalysen	106
A.4.1	Methylierungsgrad von 104 human/nicht-murinen intronischen CGIs mit zusätzlichen Treffern im Genom (Methylom 1 und 2)	106
A.4.2	Histogramme der Methylierung der einzelnen Reads der 14 Kandidaten-CGIs (Methylom 1 und 2)	108
A.4.3	Methylierungsgrad der CGIs 14414_1_hg19 (<i>ASRGL1</i>), 15224_1_hg19 (<i>PARP11</i>) und 19100_1_hg19 (<i>PDXDC1</i>) in Monozyten	112
A.4.4	Methylierungsgrad der CGIs 14414_1_hg19 (<i>ASRGL1</i>), 15224_1_hg19 (<i>PARP11</i>) und 19100_1_hg19 (<i>PDXDC1</i>) in Blut	119

A.4.5	Charakteristik von R-CGIs anhand dem Expressionslevel, der genomischen Position und dem Grad der Methylierung	121
A.4.6	Methylierungsgrad in anderen Gewebetypen der 17 Retrokopie-assoziierten CGIs	123
A.4.7	Methylierungsmuster der 12 R-CGIs mittels Roche/454 GS Junior .	126
A.5	Evolutionsanalyse	130
A.6	RePCR-Programme	132
A.7	Ansätze und PCR-Programme der RT-PCR	132
A.8	Zusätzliche Plots zu den Methylierungsanalysen der Umgebung von Retrokopie-assoziierten CGIs unter Berücksichtigung von repetitiven Sequenzen	134
A.9	Rohdaten der SNaPshot-Analyse mit anschließendem Genescan	138
B	Über die Autorin	140
B.1	Lebenslauf	140
B.2	Publikationen	141
C	Erklärungen	143

Abkürzungsverzeichnis

5caC	5-Carboxycytosin
5fC	5-Formylcytosin
5hmC	5-Hydroxymethylcytosin
5mC	5-Methylcytosin
°C	Grad Celsius
Abb.	Abbildung
A	Adenin, Nukleinbase der DNA und RNA
BAM	<i>Binary Alignment Map</i>
BER	Basenexzisionsreparatur
BisDNA	Bisulfit-behandelte DNA
BMBF	Bundesministerium für Bildung und Forschung
BWA	<i>Burrows-Wheeler Alignment</i>
bp	Basenpaar
C	Cytosin, Nukleinbase der DNA und RNA
CA	<i>chromatin accessibillity, Chromatin-Zugänglichkeit</i>
cDNA	<i>complementary</i> DNA; komplementäre DNA
CDS	<i>coding sequence</i> ; kodierende Sequenz
CGI	<i>CpG island</i> ; CpG-Insel
CpG	CpG-Dinukleotid (verbunden durch eine Phosphatbrücke)
DMR	differentiell-methylierte Region
DNA	<i>deoxyribonucleic acid</i> ; Desoxyribonukleinsäure
DNase	Desoxyribonuklease
DNMTs	DNA-Methyltransferasen
dNTP	2'-Desoxyribonucleosid-5'-triphosphat

EDTA	Ethylendiamintetraacetat
emPCR	Emulsions-PCR
FA	Formamid
for-Primer	<i>forward</i> Primer
G	Guanin, Nukleinbase der DNA und RNA
g	Erdbeschleunigung 9,81 m/s ²
g	Gramm
h	<i>hour</i> ; Stunde
IC	<i>Imprinting center</i>
ICR	<i>Imprinting contol region</i> ; Imprintingkontrollregion
ID	<i>Imprinting</i> Defekt
IGV	<i>Integrative Genomic Viewer</i>
kb	Kilobasenpaar
l	Liter
LINE	<i>long interspersed nuclear elements</i>
LTR	<i>long terminal repeat elements</i>
M	mol/l
mat	maternal
Mb	Megabasenpaar
meth	methyliert
MID	<i>Multiplex identifier</i>
min	Minute
ml	Milliliter
MLID	Multilokus-Imprinting-Defekt
mRNA	<i>messenger RNA</i> ; Boten-RNA
ng	Nanogramm
NGS	<i>next generation sequencing</i> ; Hochdurchsatzsequenzierung
NK	Negativkontrolle
nm	Nanometer

NC	<i>normal control</i> ; Normalkontrolle
OD	Optische Dichte
ORF	<i>open reading frame</i> ; offenes Leseraster
pat	paternal
PCR	<i>polymerase chain reaction</i> ; Polymerasekettenreaktion
R-CGI	Retrogen-assoziiertes CGI
rev-Primer	<i>reverse</i> Primer
RNA	<i>ribonucleic acid</i> ; Ribonukleinsäure
RNase	Ribonuklease
rpm	<i>rounds per minute</i> ; Umdrehungen pro Minute
RT	Reverse Transkription
RT-PCR	PCR im Anschluss an die reverse Transkription
SAM	<i>Sequence Alignment Map</i>
SAP	Shrimp Alkaline Phosphatase
SD	<i>standard derivation</i> ; Standardabweichung
sek	Sekunde
Seq-Primer	Sequenzierprimer
SINE	<i>short interspersed nuclear elements</i>
SNP	<i>single nucleotide polymorphism</i> ; Einzelbasenaustausch
T	Thymin, Nukleinbase der DNA
TAE	Tris-Acetat-EDTA-Puffer
TDG	Thymin-DNA-Glykosylase
TI	transkriptionelle Interferenz
Tm	Annealing Temperatur
TSS	Transkriptionsstartstelle
unmeth	unmethyliert
U	Uracil, Nukleinbase der RNA
UPD	Uniparentale Disomie
UTR	untranslatierter Region

VAR Varianz

WGBS *Whole-Genome Bisulfite Sequencing*

1 Einleitung

1.1 CpG-Inseln

CpG-Inseln (CGIs) sind DNA Regionen, die im Gegensatz zum übrigen Genom einen hohen GC-Gehalt und eine hohe Frequenz an CpG-Dinukleotiden aufweisen. In der Literatur wurden CGIs erstmals von Gardiner-Garden u. Frommer (1987) beschrieben und als CpG-reiche Regionen mit einem GC-Gehalt über 50%, einer Länge von mindestens 200bp und einer Rate von beobachteten zu erwarteten CpGs (*obs./exp. ratio*) über 0,6 definiert. Diese Rate errechnet sich nach der folgenden Formel:

$$\frac{\text{Anzahl CpGs}}{\text{Anzahl Cs} * \text{Anzahl Gs}} * \text{Anzahl aller Nukleotide}(N)$$

Da einige repetitive Sequenzen (z.B. Alu-Elemente) auch die eben genannten Kriterien erfüllen, wurden CGIs durch Takai u. Jones (2002) redefiniert (Länge größer 500bp, GC-Gehalt größer 55% und eine Rate von beobachteten zu erwarteten CpGs größer 0,65). Die meisten Programme zur Vorhersage von CGIs (inklusive dem UCSC Browser) benutzen jedoch die ursprüngliche Definition von Gardiner-Garden u. Frommer (1987) und zusätzlich weitere Tools wie z.B. den *Repeatmasker*, um repetitive Sequenzen auszuschließen.

Im menschlichen Genom gibt es insgesamt 27.537 CGIs (UCSC Browser, Stand Juni 2014), die im Durchschnitt zwischen 300bp und 3.000bp lang sind. Oft liegen CGIs am 5'-Ende von Genen (Startstellen der Transkription, TSS), man kann sie jedoch auch innerhalb von Genen (intragenisch) oder zwischen Genen (intergenisch) finden. Nur ein sehr geringer Anteil aller humanen CGIs liegt ausschließlich im Intron eines Gens (Illingworth u. Bird, 2009; Fatemi u. a., 2005). Im Gegensatz zum Menschen besitzt die Maus deutlich weniger CGIs (n=15.997, UCSC Browser, Stand Juni 2014), obwohl die Anzahl der Gene im humanen und murinen Genom in etwa gleich ist (Guenet, 2005). Etwa 70% der humanen Promotoren weisen einen hohen CpG-Gehalt auf. Wenn man jedoch von der statistischen Frequenz von CG-Nukleotiden ausgeht, ist die beobachtete Anzahl von CpG-Dinukleotiden viel geringer, als die erwartete (Saxonov u. a., 2006; Nordheim u. Knippers, 2015).

Die meisten CGIs im menschlichen Genom liegen unmethyliert vor (Jones, 2012). Jedoch kann das Cytosin eines CpG-Dinukleotids mit Hilfe einer DNA-Methyltransferase zu 5-Methylcytosin (5mC) methyliert werden (siehe Abschnitt 1.2). Methylierte CGIs können durch Proteine gebunden werden, was eine Verdichtung der Chromatinstruktur zur Folge hat und die Aktivierung eines benachbarten oder überlappenden Promotors verhindert (Nordheim u. Knippers, 2015).

1.2 DNA-Methylierung

Bei der DNA-Methylierung handelt es sich um eine chemische Modifikation, die postreplikativ durch spezifische Enzyme, den sogenannten DNA-Methyltransferasen (DNMTs), erfolgt. Die DNA-Methylierung kann in vielen Organismen beobachtet werden, jedoch ist diese in Eukaryoten auf die C5-Position des Cytosins innerhalb eines CpG-Dinukleotids beschränkt. Durch Anhängen einer Methylgruppe wird das Cytosin zu 5mC. Wie auch die Modifikationen von Histonen gehört die DNA-Methylierung von Cytosin zu den epigenetischen Mechanismen, die ohne Veränderung der Nukleotidsequenz die lokale Aktivität von Chromosomen beeinflussen (Nordheim u. Knippers, 2015; Mattick u. a., 2009). In dem Modellorganismus *C. elegans* gibt es zum Beispiel keine DNA-Methyltransferase und daher wird auch kein 5mC gebildet. Die DNA-Methylierung spielt somit in diesem Organismus keine Rolle. In *Drosophila*, einem weiteren Modellorganismus, spielt die DNA-Methylierung ebenfalls nur eine geringe Rolle. Nur während der frühen Entwicklung kann 5mC im Genom nachgewiesen werden, wobei der Anteil von 5mC bei *Drosophila* sehr gering ist (Tweedie u. a., 1997).

Wie auch schon in Abschnitt 1.1 erwähnt wurde, ist die Anzahl von beobachteten CpG-Dinukleotiden im Genom geringer als statistisch erwartet. Dieses Phänomen lässt sich mit Hilfe der DNA-Methylierung erklären. Das 5mC neigt dazu, durch Desaminierung in Thymin umgewandelt zu werden. Diese Mutation wird jedoch nicht durch das DNA-Reparatursystem erkannt und somit sind CpG-Dinukleotide im Laufe der Evolution im humanen und murinen Genom unterrepräsentiert. CpG-Dinukleotide sind nicht gleichmäßig über das Genom hinweg verteilt, sondern kommen gehäuft in repetitiven Sequenzabschnitten oder in CGIs vor (siehe Abschnitt 1.1, Nordheim u. Knippers (2015); Bird (2002); Cedar u. Bergman (2009)).

Im Laufe der Entwicklung von Säugetieren ändert sich der Grad der DNA-Methylierung,

wobei die DNA-Methylierung für die embryonale Entwicklung erforderlich ist (Bird, 2002). In primordialen Keimzellen werden das paternale und das maternale Genom erst einmal vollständig demethyliert. Auf die Phase der Demethylierung folgt später in der Keimzellentwicklung die Phase der Remethylierung. Zu diesem Zeitpunkt kann eine genomische Prägung gesetzt werden (siehe Abschnitt 1.4). In der Zygote und im frühen Embryo folgt dann eine zweite Reprogrammierungswelle. Beide Genome sind zu diesem Zeitpunkt noch räumlich voneinander getrennt. Fast alle Methylierungsmuster, die bereits in der Keimbahn gesetzt wurden, werden wieder entfernt und durch neue, auf beiden Allelen identische somatische Methylierungsmuster, ersetzt. Die genomische Prägung ist vor dieser zweiten Reprogrammierungswelle geschützt. Während dieser unterschiedlichen Phasen trägt die DNA-Methylierung zu der epigenetischen Regulation von vielen Prozessen bei (Genomische Prägung, X-Inaktivierung, Stabilität des Genoms und Genregulation)(Haaf, 2006; Hackett u. Surani, 2013; Reik u. a., 2001; Seisenberger u. a., 2013).

Unterschiedliche Methylierungsmuster werden im Laufe dieser Entwicklung durch die *de novo* Methyltransferasen DNMT3A und DNMT3B erzeugt. Diese Methyltransferasen werden in undifferenzierten embryonalen Stammzellen stark exprimiert, in somatischen Gewebszellen aber nur schwach exprimiert. Nach der DNA-Replikation (vor der Mitose oder innerhalb der Meiose) befindet sich die DNA in einem hemimethylierten Zustand (nur ein Strang trägt das Methylierungsmuster). Durch eine weitere Methyltransferase, der Erhaltungsmethyltransferase DNMT1, wird der neu synthetisierte DNA-Strang methyliert. DNMT1 folgt dem Replikationskomplex, erkennt hemimethylierte DNA (mit Hilfe von UHRF1) und fügt Methylgruppen an den neu synthetisierten DNA-Strang an. An diesem Mechanismus sind jedoch auch weitere Faktoren, wie das Chromatin, beteiligt, vor allem durch chemische Modifikationen von Histonproteinen (Okano u. a., 1999; Li u. a., 1992; Holliday u. Pugh, 1996; Dean, 2014; Fuks, 2005).

5-Methylcytosin kann stufenweise weiter oxidiert werden, wodurch eine Demethylierung der DNA stattfindet. Die hierdurch entstandenen Modifikationstypen kommen jedoch in geringerer Anzahl als 5mC vor. Zu diesen Modifikationstypen gehören 5-Hydroxymethylcytosin (5hmC), 5-Formylcytosin (5fC) und 5-Carboxycytosin (5caC). Verantwortlich für diese Oxidation sind die TET-Enzyme (*ten eleven translocation*), bei denen es sich um Fe(II)-abhängige Dioxygenasen handelt. Die Funktionen dieser unterschiedlichen Modifikationstypen sind noch nicht detailliert bekannt, jedoch gibt es Hinweise, dass 5hmC die DNA-Erkennung moduliert, häufig mit transkriptionell aktiven Bereichen assoziiert wird und wohl eine bedeutende Rolle bei der DNA-Demethylierung spielt. Die anderen beiden Formen, 5fC und 5caC sind nur kurzzeitig vorhanden und weitere Funktionen

sind momentan noch nicht bekannt (Kohli u. Zhang, 2013; Pastor u. a., 2013; Hackett u. Surani, 2013; Schübeler, 2015). Desweiteren können 5fC und 5caC durch ein Enzym, die so genannte Thymin-DNA-Glykosylase (TDG), entfernt (Exzision) und anschließend durch die Basenexzisionsreparatur (BER) durch unmodifizierte Cytosine ersetzt werden (Krokan u. Bjørås, 2013).

Eine weitere mögliche Funktion der TET-Enzyme liegt in der DNA-Demethylierung. Hierbei konnten sowohl aktive als auch passive Mechanismen beobachtet werden (Pastor u. a., 2013; Wu u. Zhang, 2014; Jin u. a., 2014). TET-Enzyme sind möglicherweise für die Aufrechterhaltung des unmethylierten Zustands von CGIs verantwortlich (aktive Demethylierung). Jin u. a. (2014) konnten in HEK293T-Zellen zeigen, dass eine Anreicherung von 5hmC am den Rändern von unmethylierten CGIs auf TET1 zurückzuführen ist. Eine Ausbreitung der DNA-Methylierung in diese hypomethylierten CGIs ist somit nicht möglich.

Die DNA-Methylierung kann auch die Bindung von Transkriptionsfaktoren (z.B. SP1) und Isolatorproteinen (z.B. CTCF) beeinflussen. Liegen solche Bindestellen methyliert vor, können die DNA-bindenden Proteine sich nicht an die DNA anlagern und die Transkription kann nicht stattfinden. Gene können also durch DNA-Methylierung dauerhaft ausgeschaltet werden (Robertson, 2005).

Es gibt unterschiedliche Methoden die DNA-Methylierung zu analysieren wobei die Behandlung von DNA mit Bisulfit eine weit verbreitete Methode ist. Hierbei wird der Schutz von methylierten Cytosinen vor einer chemischen Umwandlung in Uracil benutzt. Es werden also nur unmethylierte Cytosine in Uracil umgewandelt. Während der anschließenden Polymerasekettenreaktion (*polymerase chain reaction*, PCR) wird das Uracil durch Thymin ersetzt (Frommer u. a., 1992). Die spezifischen Sequenzunterschiede können dann anschließend z.B. mittels Klonierung und anschließender Sequenzierung oder heutzutage auch mittels quantitativer Hochdurchsatzsequenzierung (*next generation sequencing*, NGS) analysiert werden.

1.3 Histonmodifikationen

Die Modifikationen von Histonen sind ebenfalls ein wichtiger epigenetischer Prozess im humanen Genom. Histone sind basische Proteine, die eine positive Ladung tragen. Sie sind im Zellkern aller Tier- und Pflanzenzellen zu finden und Bestandteil des Chromatins.

Insgesamt sind fünf Hauptklassen von Histonen bekannt: H1, H2A, H2B, H3 und H4. Histone sind für die Verpackung der DNA zuständig. In Eukaryoten passiert dies in den Chromosomen. Die kleinste Einheit der Verpackung sind die Nukleosome. Diese bestehen aus der negativ geladenen DNA, die um ein Histonoktamer (jeweils zwei Kopien der Histone H2A, H2B, H3 und H4) gewickelt ist. Histone besitzen terminale Bereiche, sogenannte *histon tails*, die basische Aminosäuren besitzen und aus dem Nukleosom herausragen. Diese terminalen Bereiche können durch verschiedene Enzyme modifiziert werden. Zu den bekanntesten Arten der Histonmodifikation gehören neben der Methylierung, die Phosphorylierung, die Ubiquitinierung und die Acetylierung (Nordheim u. Knippers, 2015; Li, 2002).

Die Übertragung der Modifikationsmuster erfolgt durch spezifische Enzyme, den Transferasen (Histon-Methyltransferasen und Histon-Acetyltransferasen). Das Entfernen von Mustern geschieht durch Histon-Demethylasen und Histon-Deacetylasen. Zu den prominentesten Modifikationstypen gehören neben dem Verschließen des Chromatins (H2AK119ub), das Öffnen des Chromatins und damit die Aktivierung der Transkription (H3K4me1, H3K4me3, H3K9ac und H3K27ac) und auch die Blockade der Transkription (H3K9me3, H3K27me3, H4K20me1, H4K20me3). Der Name der jeweiligen Histonmodifikation setzt sich aus der Histonklasse (z.B. H1 oder H2A), der Aminosäure (AS) die modifiziert wird (z.B. K119 oder K4, wobei K für die AS Lysin steht) und der Art der Modifikation (z.B. me3, das für Trimethylierung steht) zusammen (Jenuwein u. Allis, 2001; Li, 2002; Bannister u. Kouzarides, 2011; Zhou u. a., 2011; Greer u. Shi, 2012).

Die Histonmethylierung spielt eine bedeutende Rolle in vielen biologischen Prozessen, wie z.B. der Regulation des Zellzyklus, der Stress-Antwort, der Entwicklung, der Differenzierung und ist auch bei verschiedenen Krankheiten und dem Alterungsprozess von großer Bedeutung (Greer u. Shi, 2012). Desweiteren spielen Histonmodifikationen und DNA-Methylierung zusammen eine zentrale Rolle für die Steuerung epigenetischer Prozesse wie z.B. der genomischen Prägung (siehe nächster Abschnitt 1.4).

1.4 Genomische Prägung

Genomische Prägung (*genomic imprinting*) ist ein epigenetischer Prozess, der die vererbte Aktivierung bzw. Inaktivierung von Genloci in Abhängigkeit ihrer elterlichen Herkunft beschreibt. In somatischen Zellen ist somit entweder nur das paternale oder nur

das maternale Allel eines Gens aktiv (monoallelische Expression).

Wie bereits in Abschnitt 1.2 beschrieben, werden während der Embryonalentwicklung sowohl das paternale als auch das maternale Genom demethyliert; Die Demethylierung des maternalen Genoms findet zeitlich etwas später statt. Anschließend erfolgt eine Remethylierung der beiden Genome. Da die elterliche Prägung bereits in der Keimbahn gesetzt wurde, sind geprägte Regionen vor der De- und Remethylierung geschützt (Reik u. Walter, 2001; Ideraabdullah u. a., 2008; Horsthemke, 2010).

In primordialen Keimzellen erfolgt ebenfalls eine vollständige Demethylierung, die auch geprägte Regionen betrifft. Die Prägung wird in diesem Fall gelöscht (*imprint erasure*). Im anschließenden Schritt werden dann die dem Geschlecht entsprechenden Methylierungsmuster u.a. neu etabliert (*imprint establishment*). Die elterliche Prägung wird in allen folgenden Zellteilungen aufrechterhalten (*imprint maintenance*) (Reik u. Walter, 2001; Edwards u. Ferguson-Smith, 2007; Horsthemke u. Buiting, 2008; Hanna u. Kelsey, 2014; Barlow u. Bartolomei, 2014; Macdonald u. Mann, 2014).

Die genomische Prägung wurde erstmals 1984 von Surani u. a. (1984) und McGrath u. Solter (1984) beschrieben. Anhand von Transplantationsexperimenten von Zellkernen in Mäusen konnte gezeigt werden, dass beide elterlichen Genome essentiell für die Embryonalentwicklung sind. Desweiteren wurde gezeigt, dass die Funktionalität von beiden elterlichen Genomen nicht identisch ist. So zeigten Embryos mit zwei paternalen Zellkernen, also ausschließlich mit paternalem Genom (androgenetisch), extraembryonale Strukturen, jedoch kaum embryonales Gewebe. Im Gegensatz hierzu zeigten Embryos mit zwei maternalen Zellkernen (gynogenetisch bzw. parthenogenetisch) hauptsächlich embryonales Gewebe jedoch kaum extraembryonales Gewebe. Weder in Embryonen mit einem diploiden maternalen Chromosomensatz noch in Embryonen mit einem diploiden paternalen Chromosomensatz konnte eine normale Embryonalentwicklung beobachtet werden. Die Embryonen starben früh während der Tragzeit ab. Die eben beschriebenen Beobachtungen von uniparentalen Genomen wurden in ähnlicher Form auch beim Menschen gemacht. Ein androgenetisches Genom führt zur Entwicklung einer Blasenmole, die keine embryonalen Strukturen aufweist, sondern aus extraembryonalem Gewebe besteht. Im Gegensatz hierzu führt ein gynogenetisches Genom zur Entwicklung eines Ovarialteratoms mit embryonalem Gewebe aller drei Keimblätter, wobei jedoch extraembryonales Gewebe fehlt (Hall, 1990).

Viele geprägte Gene sind bereits in der Plazenta und/oder im Fötus aktiv und regulieren die Ressourcenverteilung zwischen Mutter und Kind (Haaf, 2006). Paternal exprimierte Gene können z.B. das Wachstum des Föten fördern, wohingegen maternal exprimierte Gene das Wachstum hemmen. Ein Beispiel hierfür ist das von Barlow u. a. (1991) beschriebene

erste murine geprägte Gen *Igf2r* (*insulin-like growth factor 2 receptor*). Das *Igf2r*-Gen wird nur vom maternalen Allel exprimiert und kodiert für den Igf2-Rezeptor. Im Gegensatz hierzu wird das *Igf2*-Gen, das für einen Wachstumsfaktor kodiert, nur vom paternalen Allel exprimiert. Die maternale Expression von *Igf2r* bewirkt, dass der Wachstumsfaktor Igf2 abgefangen und das Wachstum somit gehemmt wird. Dieses Beispiel unterstützt auch die Geschlechterkonflikthypothese die von Moore u. Haig (1991) publiziert wurde. Die Hypothese beruht darauf, dass Frauen ihre Gene nur durch eigene Schwangerschaften vererben können und den Verbrauch ihrer Ressourcen für weitere Schwangerschaften niedrig halten wollen. Im Gegensatz hierzu kann der Mann seine Gene über mehrere bzw. verschiedene Frauen weitervererben. Wie im oben genannten Beispiel (*Igf2r* und *Igf2*) fördern paternal exprimierte Gene das Wachstum des Fötus, wohingegen maternal exprimierte Gene das Wachstum hemmen (Moore u. Haig, 1991; Haig, 1997).

Häufig liegen im Promotorbereich von geprägten Genen CGIs (siehe Abschnitt 1.1). Diese überlappen in den meisten Fällen mit differentiell methylierten Regionen (DMRs). DMRs zeichnen sich durch ein Methylierungsniveau von ca. 50% aus, wobei ein Allel methyliert und das andere unmethyliert ist. Eine Expression des Gens findet somit nur auf dem unmethylierten Allel statt, so dass man von einer monoallelischen und elternspezifischen Expression des Gens spricht (Bartolomei u. Ferguson-Smith, 2011).

Eine elternspezifische Expression von geprägten Genen kann manchmal auch nur in einem bestimmten Gewebe oder aber auch zu einem bestimmten Zeitpunkt während der Entwicklung stattfinden (Prickett u. Oakey, 2012). Ein Beispiel hierfür ist z.B. das humane *UBE3A*-Gen. Es wird in fast allen Geweben biallelisch exprimiert außer im Gehirn, in dem es monoallelisch vom maternalen Allel exprimiert wird (Rougeulle u. a., 1997).

Heutzutage sind ca. 100 humane geprägte Gene (www.geneimprint.com, www.igc.otago.ac.nz) und ca. 150 murine geprägte Gene (www.geneimprint.com, www.igc.otago.ac.nz, www.mousebook.com) bekannt, von denen die Mehrheit in beiden Organismen vorhanden ist. Geprägte Gene sind zufällig im Genom verteilt, liegen oft aber gehäuft in chromosomalen Regionen, so genannten Clustern. Bekannte Cluster sind z.B. auf den Chromosomen 6, 7, 11, 14, 15 und 20 zu finden. Die meisten dieser Cluster sind mit *Imprinting*erkrankungen assoziiert und werden durch in *cis*-agierenden *Imprinting*kontrollregionen (*imprinting control region*, ICR oder auch *imprinting center*, IC, genannt) reguliert (Maher u. Reik, 2000; Reik u. Walter, 2001; Horsthemke, 2010; Ferguson-Smith, 2011; Bartolomei u. Ferguson-Smith, 2011; Horsthemke, 2014).

Unterschiedliche Krankheiten, z.B. das Angelman-, das Prader-Willi-, das Beckwith-Wiedemann- und das Silver-Russel-Syndrom, lassen sich auf Störungen der genomischen

Prägung zurückführen. Diesen sogenannten *Imprinting*erkrankungen liegen verschiedene molekulargenetische Mechanismen zugrunde, wie Deletionen, uniparentalen Disomien (UPD), Genmutationen, Mikrodeletionen im IC oder Duplikationen geprägter Gene. Zwischen den einzelnen *Imprinting*defekten (IDs) wird im Allgemeinen zwischen primären und sekundären IDs unterschieden. Primäre IDs sind Defekte ohne Veränderung der Nukleotidsequenz selbst, sondern nur die Methylierung ist betroffen. Im Gegensatz hierzu beruhen sekundäre IDs auf Sequenzveränderung, die zur Änderung der Methylierung führen.

Durch Deletionen (z.B. bei Angelman- und Prader-Willi-Syndrom) kann der Chromosomenabschnitt mit der exprimierten Genkopie verloren gehen. Diese Deletionen entstehen meist *de novo*, womit ein Wiederholungsrisiko für weitere Schwangerschaften gering ist. Desweiteren können *Imprinting*erkrankungen durch UPDs entstehen. Hierbei stammen zwei homologe Chromosomen entweder nur von der Mutter oder nur vom Vater. Gelegentlich kann durch eine Genmutation ein Gen, das der genomischen Prägung unterliegt, seine Funktion verlieren. Eine Mikrodeletion im IC kann ebenfalls zu einer *Imprinting*erkrankungen führen, so dass es zum Verlust der normalen Methylierung kommt. Eine Duplikation eines geprägten Gens kann auch zu einer Krankheit führen, da es bei dem geprägten Gen auf eine Balance zwischen mütterlichen und väterlichen Genen ankommt, wie es beispielsweise bei Wachstumsförderern und Wachstumshemmern der Fall ist (Beckwith-Wiedemann-Syndrom). In manchen Patienten ist jedoch nicht nur ein geprägter Locus betroffen. Bei diesen Patienten spricht man dann von einem Multilocus-*Imprinting*-Defekt (MLID) (Horsthemke, 2014; Ferguson-Smith, 2011; Bartolomei u. Ferguson-Smith, 2011; Barlow u. Bartolomei, 2014).

In den letzten Jahren wurden viele genomweite Studien zur DNA-Methylierung und zur genomischen Prägung publiziert. Unter diesen waren neben theoretischen Ansätzen mit rechenintensiven Vorhersagemodellen, wie veröffentlicht von Luedi u. a. (2005) und Laird (2010), auch experimentelle Ansätze mittels Microarrays oder quantitativer Hochdurchsatzsequenzierung, z.B. publiziert von Lister u. a. (2010). Ein anderer Fokus von Studien über genomische Prägung liegt in der Evolution von CGIs in unterschiedlichen Säugetieren. Bis jetzt ist über molekulare und evolutionäre Mechanismen, die im Laufe der Evolution entstanden sind, wenig bekannt. Zu diesen Mechanismen gehört auch die genomische Prägung (Suzuki u. a., 2011). Ein weiterer Schwerpunkt der Forschung über genomische Prägung beschäftigt sich mit der Suche nach elternspezifischen DMRs, wie z.B. von Court u. a. (2014) publiziert wurde.

1.5 Retrotransposition

Neue Gene entstehen hauptsächlich durch Genduplikationen, welche somit für die Evolution von großer Bedeutung sind (Kaessmann, 2010). Schon Ohno (Ohno, 1970) schrieb 1970, dass die Genduplikation eine Hauptkraft der Evolution sei. Es gibt zwei grundlegende Mechanismen, durch die Genduplikationen entstehen können. Die segmentale Duplikation beschreibt einen Mechanismus der Genduplikation, bei der exprimierte Genkopien mit Intron-Exon-Struktur und regulatorischen Elementen entstehen, welche häufig in unmittelbarer Nähe liegen. Der zweite Mechanismus ist die Retrotransposition oder auch Retroduplikation. Hierbei entstehen Genkopien ohne Introns, Promotoren und regulatorischen Elementen, die für gewöhnlich nicht exprimiert werden und häufig auf anderen Chromosomen liegen als ihr ancestrales Gen (Zhang, 2003; Kaessmann u. a., 2009; Kaessmann, 2010).

Unter Retrotransposition versteht man die Integration einer revers-transkribierten mRNA (cDNA) in eine andere genomische Region, wobei die reverse Transkriptase eine bedeutende Rolle spielt. Dieses Enzym kann durch unterschiedliche transponierbare Elemente in unterschiedlichen Organismen kodiert werden (Nordheim u. Knippers, 2015). Im Humangenom gehören die LINEs (*long interspersed nuclear elements*) zu den wichtigsten transponierbaren Elementen, die die Retrotransposition von Genen ermöglichen. In höheren Eukaryoten sind bis zu 100.000 Kopien dieser Elemente zu finden. Daher können Retrotranspositionen in Säugetieren auch besonders häufig beobachtet werden. LINEs kodieren für eine reverse Transkriptase mit endonukleolytischer Aktivität, die neben der eigenen RNA auch polyadenylierte mRNA binden kann. Die mRNA der ancestralen Gene wird revers transkribiert und die cDNA an eine neue Stelle im Genom wieder eingefügt (Esnault u. a., 2000; Wei u. a., 2001; Torrents u. a., 2003; Vinckenbosch u. a., 2006; McCole u. a., 2011).

Über einen langen Zeitraum war eine mögliche Funktion von Retrokopien bzw. die Entwicklung zu exprimierten Retrogenen unbekannt, da Retrokopien weder Introns, Promotoren noch regulatorischen Elemente an ihrer Integrationsstelle im Genom aufwiesen. Sie wurden daher auch prozessierte Pseudogene genannt (Mighell u. a., 2000; Kaessmann u. a., 2009; Kaessmann, 2010). Vinckenbosch u. a. (2006) haben in ihrer Studie Retrokopien auf Transkription untersucht und herausgefunden, dass im menschlichen Genom mehr als 1.000 Retrokopien transkribiert werden. Mehr als 10% davon haben sich im Laufe der Evolution zu *bona fide* Genen entwickelt.

Inzwischen sind mehrere unterschiedliche Mechanismen bekannt, wie sich Retrokopien zu funktionellen Genen (Retrogenen) entwickeln können. Zum einen können Retrokopien von

vorhandenen regulatorischen Elementen, wie z.B. Promotoren, an ihrer neuen Integrationsstelle profitieren. Dies kann z.B. der Fall sein, wenn eine Retrokopie in einem Intron eines anderen Gens liegt. Die Retrokopie kann dann zusammen mit dem eigentlichen Gen als Fusionstranskript transkribiert werden (Kaessmann u. a., 2009). Auch ist es möglich, dass die Transkription von Retrokopien durch entfernte CpG-reiche Sequenzen, die als Promotor dienen, gesteuert wird. Durch natürliche Selektion können sich aus diesen CpG-reichen Sequenzen vollwertige Promotoren entwickeln (Vinckenbosch u. a., 2006; Kaessmann u. a., 2009). Ein Promotor kann auch schon in einer Retrokopie vorhanden sein, wenn das anze-strale Gen zum Beispiel mehrere alternative Transkriptionsstartstellen hat (Kaessmann u. a., 2009).

Bekannt sind auch Zusammenhänge zwischen Retrotransposition und genomischer Prägung. In der Maus sind fünf geprägte Gene (*Mcts2*, *Nap1l5*, *U2af1-rs1*, *Inpp5f_v2* und *Peg12*) bekannt, die durch Retrotransposition entstanden sind (Wood u. a., 2007; Cowley u. Oakey, 2010; McCole u. a., 2011). Auch im Humangenom ist dieses Phänomen bei geprägten Regionen beobachtet worden, wie z.B. bei: *MAGEL2*, *MKRN3*, *NPAP1*, *NDN* und *RB1* (Kanber u. a., 2009b; Cowley u. Oakey, 2010; Zhang u. a., 2011; Neumann u. a., 2014).

1.6 Das geprägte *RB1*-Gen

Bei dem *RB1*-Gen handelt es sich um ein Tumorrepressorgen (NM_000321), das das nukleare Phosphoprotein pRb codiert (Friend u. a., 1986; Lee u. a., 1986). Mutationen in beiden Allelen des Gens führen in früher Kindheit zu einem bösartigen Tumor der Netzhaut des Auges. Laut Kivelä (2009) sind Retinoblastome die häufigsten primären Augentumoren des Kleinkindes.

Das humane *RB1*-Gen liegt auf Chromosom 13 und ist geprägt (Kanber u. a., 2009b). Die genomische Prägung dieses Gens beruht auf einem differentiell methyliertem CGI (CpG85; paternal unmethyliert, maternal methyliert) in Intron 2 des *RB1*-Gens. CpG85 ist Teil einer durch Retrotransposition eingefügten Sequenz des *PPP1R26*-Gens von Chromosom 9 und hat sich zu einem alternativen *RB1* Promotor entwickelt (Kanber u. a., 2013) (siehe Abb. 1.1).

Erstmals beschrieben wurde dieses paternal exprimierte alternative *RB1*-Transkript von Kanber u. a. (2013). Desweiteren konnten Kanber u. a. (2013) in ihrer Arbeit eine Expressionsimbalance (*skewing*) zugunsten des maternalen Transkripts im Verhältnis von

3:1 feststellen. Um einen möglichen Zusammenhang zwischen dem alternativen *RB1*-Transkript und dem verschobenen Transkriptverhältnisses zu klären, führten Kanber u. a. (2013) Zellkultur-Experimente durch. Diese Experimente zeigten, dass nach einer Demethylierung des CpG85 das maternale Allel aktiviert und damit das alternative *RB1*-Transkript biallelisch exprimiert wurde. Diese gleichzeitige Expression beider Allele führte desweiteren zu einer geringer verschobenen Expression des regulären *RB1*-Transkripts. Aufgrund dieses Ergebnisses schlossen die Autoren darauf, dass die Expression des alternativen Transkripts zur Reduktion der Expression des regulären *RB1*-Transkripts führt und damit von dem Methylierungsstatus des CpG85 abhängig ist. Als ein mögliches Erklärungsmodell geben Kanber u. a. (2013) die so genannte transkriptionelle Interferenz (TI) an.

Unter TI versteht man eine bestimmte Form der Genregulation, bei der die Transkription an Promotor A einen direkten Einfluss (oder in *cis*) auf die Transkription an Promotor B hat. Dabei können zum Beispiel beide Promotoren (im Beispiel des geprägten *RB1*-Gens wären das der reguläre *RB1*-Promotor CpG106 und der alternative *RB1*-Promoter CpG85) gleichzeitig Transkriptionskomplexe beiden Promotoren binden, die sich dann aber gegenseitig blockieren, was u.a. durch eine unterschiedliche Geschwindigkeit der Transkription geschehen kann (Shearwin u. a., 2005).

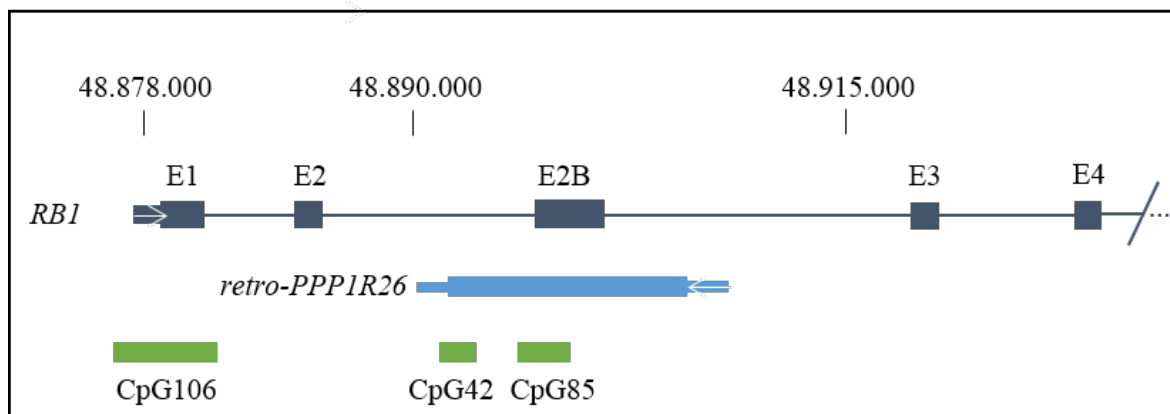


Abbildung 1.1: Schematische Darstellung der *RB1*-Gens im Humangenom

Die Abbildung zeigt die ersten vier Exons des *RB1*-Gens. RefSeq-Gene sind in dunkelblau, Retrogene in hellblau und CGIs in hellgrün dargestellt. Weiße Pfeile zeigen die Richtung der Transkription an. Die Zahlen spiegeln die Position auf Chromosom 13 wieder (hg19). Die Abbildung ist nicht maßstabsgerecht gezeichnet.

Andere Studien, u.a. Nakabayashi u. a. (2011), konnten die allelische Methylierung des CpG85 durch genomweite Microarray-Analysen (*Infinium HumanMethylation27 BeadChip*,

Illumina) bestätigen. Im Gegensatz zum humanen *RB1*-Gen ist das murine *Rb1*-Gen nicht geprägt und besitzt auch kein Retrotransposon (Steenpass u. a., 2013).

1.7 Zielsetzung

Ausgangspunkt dieser Arbeit ist das humane *RB1*-Gen, welches geprägt ist. Durch Integration einer Retrokopie, welche das CpG85 beinhaltet, erlangte das *RB1*-Gen seine genomische Prägung. Im Laufe der Evolution entwickelte sich aus dem CpG85 ein alternativer, differentiell methylierter Promotor, der für die geprägte Expression des *RB1*-Gens verantwortlich ist. Das murine *Rb1*-Gen besitzt keine Retrokopie und ist daher auch nicht geprägt (Kamber u. a., 2009a, 2013; Steenpass u. a., 2013).

Ziel des ersten Teils dieser Arbeit ist es mit Hilfe von bioinformatischen Methoden humane intronische CGIs zu detektieren, die nicht in der Maus vorkommen. Da die Retrotransposition des ancestralen *PPP1R26*-Gens sich mehrfach im menschlichen Genom ereignet hat, soll auch diese Eigenschaft von humanen intronischen CGIs überprüft werden. Wenn möglich soll der evolutionäre Zeitpunkt der Entstehung der CGIs durch bioinformatische Methoden eingegrenzt werden.

Anschließend soll die Methylierung von human/nicht-murin intronischen CGIs in zwei Monozyten-Methylomen und bereits veröffentlichten Methylomen anderer Gewebe untersucht werden, um differentiell methylierte CGIs zu identifizieren, die einen Methylierungsgrad zwischen 20% und 80% aufweisen.

Im zweiten Schritt soll eine mögliche differentielle Methylierung von CGIs experimentell durch quantitative Methylierungsanalysen mittels Hochdurchsatz-Bisulfitsequenzierung auf dem Roche/454 GS Junior verifiziert werden, um neue geprägte Loci zu identifizieren.

Im zweiten Teil der hier vorliegenden Arbeit soll ein neuer bioinformatischer Ansatz zum Auffinden von geprägten Genen untersucht werden. Hierzu sollen alle CGIs identifiziert werden, die mit einer Retrokopie überlappen, unabhängig von ihrer Lage und ihrer evolutionären Entstehung im menschlichen Genom. Wie im ersten Teil der vorliegenden Arbeit, soll anschließend die Methylierung aller Retrokopie-assoziierten CGIs in bereits publizierten Methylomen untersucht werden. CGIs, die eine differentielle Methylierung aufweisen, sollen anschließend, wie auch im ersten Teil durch quantitative Methylierungsanalysen, experimentell verifiziert werden, um neue geprägte Loci zu identifizieren.

2 Material und Methoden

2.1 Daten

Als Grundlage dieser Arbeit wurden die folgenden Datensätze untersucht:

Human

<i>CGI sequence track</i> (04. September 2014)	UCSC Browser	GRCh37/hg19	27.537 CGIs
<i>CGI information track</i> (04. September 2014)	UCSC Browser	GRCh37/hg19	27.537 CGIs
<i>retroposed gene track</i> (20. Oktober 2014)	UCSC Browser	GRCh37/hg19	13.173 Retrokopien
<i>RefSeq gene track</i> (16. Dezember 2014)	UCSC Browser	GRCh37/hg19	49.451 Gene
<i>RepeatMasker</i> (16. Dezember 2015)	UCSC Browser	GRCh37/hg19	5.189.085 repetitive Sequenzen
Transkriptionsstart und -ende (04. September 2014)	HUGO	GRCh37/hg19	

Murin

<i>CGI sequence track</i> (04. September 2014)	UCSC Browser	NCBI/mm10	15.997 CGIs
<i>CGI information track</i> (04. September 2014)	UCSC Browser	NCBI/mm10	15.997 CGIs

Humane und murine proteinkodierende Regionen

Consensus CDS (CCDS) Projekt (04. September 2014)	NCBI CCDS Datenbank	hg19/mm10	14.990 CCDS
--	---------------------	-----------	-------------

Methylierungsdaten

Methylom 1	ENA	PRJEB5800 ¹
Methylom 2	EGA	EGAS000001000719 ¹
18 publizierte Methylome	GEO NCBI	GSE46644 ²
2 publizierte Methylome	JGA	S000000000006 ³

¹ veröffentlicht in Rademacher u. a. (2014)

² veröffentlicht in Ziller u. a. (2013)

³ veröffentlicht in Okae u. a. (2014)

Die URLs der verwendeten Datensätze sind im Anhang A.1 aufgelistet. Das Datum in Klammern gibt das Datum des Downloads an. Die Analysen in dieser Arbeit beschränken sich auf die Chromosomen 1 bis 22 und Chromosom X, das Y-Chromosom wurde in dieser Arbeit nicht berücksichtigt. Für das Zusammenfügen von Informationen und die Auswertung wurden in der vorliegenden Arbeit die Programmiersprache Perl und die Statistik-Software R verwendet. Die URLs aller verwendeter Programme sind im Anhang A.1 aufgelistet. Die Perl-Skripte befinden sich auf der beigefügten Daten-CD.

2.1.1 CGI-Daten

Die CGI-Daten beinhalten alle humanen bzw. murinen CGIs, die die Kriterien von Gardiner-Garden u. Frommer (1987) erfüllen. Für jeden Organismus gibt es einen Datensatz in dem jede Zeile für ein CGI steht. Das folgende Beispiel zeigt einen Ausschnitt aus dem humanen CGI Datensatz:

Beispiel 2.1: CGI-Daten

```
<1> <2>   <3>   <4>       <5> <6> <7> <8> <9>   <10> <11>
...
586 chr1 135124 135563 CpG:30 439 30 295 13.7 67.2 0.64
587 chr1 327790 328229 CpG:29 439 29 295 13.2 67.2 0.62
...
```

Die erste Spalte ist nur zum Indizieren innerhalb des UCSC Browsers und wird somit in dieser Arbeit nicht beachtet. In den Spalten zwei bis vier steht die genaue Position eines CGIs. Neben dem Chromosom sind die genauen Start- und Endpositionen gespeichert. In der Spalte fünf befindet sich der Name des CGIs innerhalb des UCSC Browsers. Dieser Name ist jedoch nicht eindeutig, sondern setzt sich nur aus „CpG“ und der Anzahl an CpG-Dinukleotiden zusammen. Die letzten sechs Spalten charakterisieren ein CGI. Neben der Länge in Basenpaaren, der Anzahl von CpG-Dinukleotiden, der Anzahl von C- und G-Nukleotiden wird auch der Prozentsatz von CpG-Dinukleotiden, der GC-Gehalt und die *obs./exp. ratio* angegeben. Bei der dazugehörigen Sequenzdatei handelt es sich um eine Multi-FASTA-Datei, in der für jedes CGI ein FASTA-Eintrag mit Kopfzeile (beginnt mit > und enthält Informationen wie zum Beispiel das Referenzgenom und die Position im Genom) und der Sequenz gespeichert ist. Das folgende Beispiel zeigt einen FASTA-Eintrag für ein CGI, wobei nur die ersten drei Zeilen der Sequenz gezeigt werden.

Beispiel 2.2: CGI-Sequenz

```
...
>hg19_cpgIslandExt_CpG: 85 range=chr13:48892636-48893857 5'pad=0 3'pad=0 strand=+ repeatMasking=none
CGTCATGCGAGTGTGCGGGCCCTTGGCTGCGCCCTCCCTCCCGGGA
CCTCGCTGGCTTTGCCCTGGACGGGCAGGTCCGCCCTCATGGCCGGGC
ACGGCTCACGCCCTGGCTGTGGTGGCATCCTGGCCACCATCTTTCAACCAC
...
```

2.1.2 CCDS-Daten

Der CCDS-Datensatz wurde ebenfalls vom UCSC Browser heruntergeladen. Bei dem CCDS Projekt handelt es sich um ein internationales Konsortium, dessen Ziel es ist, einen Datensatz von proteinkodierenden Regionen zu bekommen, die für unterschiedliche Organismen einheitlich annotiert und von hoher Qualität sind (Pruitt u. a., 2009). Das folgende Beispiel zeigt den Aufbau von zwei humanen CCDS-Einträgen.

Beispiel 2.3: CCDS-Daten

```
<1> <2>   <3> <4> <5>   <6>   <7>   <8>   <9> <10>           <11>           <12> <13> <14> <15>   <16>   <17>
...
592 CCDS6.1 chr1 + 948953 949858 948953 949858 2 948953,949363, 948956,949858, 0 cml1 cml1 0,0, NM_005101 ISG15
592 CCDS3257.1 chr1 -1007196 1007946 1007196 1007946 1 1007196 1007946 0 cml1 cml1 0 NM_001205252 RNF223
...
```

Die erste Spalte dient wiederum der Indizierung innerhalb des UCSC Browsers (vgl. 2.1.1). Die folgenden Spalten beinhalten Name, Chromosom, Strang, Transkriptionsstart, Transkriptionsende, Start der kodierenden Region, Ende der kodierenden Region, Anzahl der Exons, Startpositionen der Exons (durch Kommata separiert), Endpositionen der Exons (ebenfalls durch Kommata separiert), ein USCS interner Score, Vollständigkeit von CDS-Start und CDS-Ende, Anzahl der Exonframes, RefSeq-ID und den Gennamen. Die letzten zwei Spalten wurden durch verlinken innerhalb des UCSC table Browsers mit RefSeq-Informationen hinzugefügt.

2.1.3 Retrokopie-Daten

Die Retrokopie-Daten wurden ebenfalls über den UCSC Browser heruntergeladen. Wie auch bei den anderen Daten dient auch hier die erste Spalte der Indizierung innerhalb des Browsers (siehe Beispiel). Die folgenden drei Spalten beschreiben die Position im Genom (Chromosom, Start und Ende). Anschließend folgen der Name der Retrokopie, der Score der Retrokopie (wird berechnet aus der Basenübereinstimmung von Gen und Retrokopie) und der Strang auf dem es liegt. Die achte Spalte wird vom UCSC Browser als *label* beschrieben und soll eigentlich genauere Informationen über die Retrokopie geben (1=Pseudogen,

-1=kein Pseudogen, -2=exprimiertes Retrogen), die aber laut UCSC Browser veraltet sind und somit in dieser Arbeit nicht berücksichtigt werden. In den folgenden fünf Spalten stehen Informationen über die proteinkodierende Region, mit der die Retrokopie überlappt. Hierbei handelt es sich um den Namen, das Chromosom, die Startposition, die Endposition und den Gennamen. Die letzte Spalte gibt an, ob die Retrokopie exprimiert wird oder nicht.

Beispiel 2.4: Retrokopie-Daten

```
<1> <2>   <3>   <4>   <5>           <6><7><8> <9>   <10>   <11>   <12>   <13>   <14>
...
590 chr1 674434 675008 NM_001743.4-7 827 + 1 NM_001743 chr2 47387220 47403740 CALM2 expressed weak
591 chr1 808688 809739 NM_080386.3-1 612 + -2 NM_080386 chr2 132233579 132240507 TUBA3D expressed weak
...
```

Im Gegensatz zu den Informationen der Spalte *label* handelt es sich bei diesen Informationen laut Email-Korrespondenz mit dem UCSC Browser-Team um aktuelle und verifizierte Daten. Daher wurden diese Informationen auch innerhalb dieser Arbeit zur Klassifizierung von Retrokopien benutzt.

2.1.4 Gen-Daten (RefSeq)

Die humanen Gen-Daten wurden ebenfalls über den UCSC Browser heruntergeladen. Als Datensatz wurde innerhalb dieser Arbeit der RefSeq-Datensatz ausgewählt. Wie auch bei den anderen Daten dient auch hier die erste Spalte der Indizierung innerhalb des Browsers (siehe Beispiel).

Beispiel 2.5: Gen-Daten

```
<1> <2>   <3> <4>   <5>   <6>   <7>   <8>   <9> <10>           <11>           <12> <13> <14> <15>   <16>
...
1 NM_013943 chr1 + 25071759 25170815 25072044 25167428 6 25071759,25124232... 25072116,25124342... 0 CLIC4 cpl1 cpl1 0,0,2,2,1,0,
...
```

Die zweite Spalte gibt den Namen des Gens an, wobei es sich um die *transcription_ID* handelt. Die folgenden sechs Spalten geben Informationen über Chromosom, Strang, Transkriptionsstart, Transkriptionsende, Start und Ende der kodierenden Sequenz des Gens. In der neunten Spalte steht die Anzahl der Exons des Gens und in den beiden folgenden Spalten sind die Startpositionen und Endpositionen der Exons angegeben. Die folgende Zeile gibt einen internen Score innerhalb des UCSC Browsers an, der nicht weiter erläutert wird. Die Spalte 13 gibt einen zweiten Namen für das jeweilige Gen an, hierbei handelt es sich um die *gene_ID*. Spalten 14 und 15 geben an, ob der CDS-Start bzw. das CDS-Ende vollständig sind. Die letzte Spalte beschreibt die Exonframes.

2.1.5 Methylo-Daten

Die Methylo-Daten (Methylo 1 und Methylo 2) liegen als an das Referenzgenom ausgerichtete (*mapped*) Reads im BAM-Format vor. Bei BAM-Dateien handelt es sich um eine komprimierte binäre Version des SAM-Formates (Li u. a., 2009). Informationen über jeden einzelnen Read (Start, Ende, Länge, Sequenzfolge, Qualität der Base, ...) und über die Region des Referenzgenoms, mit dem dieser Read übereinstimmt, sind in diesem Format speicherplatzsparend verschlüsselt. Zusätzlich zu den Reads liegen noch weitere Informationen über den Grad der Methylierung im BED-Format vor. Dieses Format hat drei vorgeschriebene Spalten für jeden Eintrag: Chromosom, Startposition und Endposition. Neun zusätzliche Spalten können optional hinzugefügt werden. Das folgende Beispiel zeigt einen Ausschnitt einer BED-Datei eines Methylo-Datensatzes publiziert von Ziller u. a. (2013) (GSE46644).

Beispiel 2.6: Methylo-Daten (1)

```
<1> <2> <3> <4> <5> <6>
...
chr1 10468 10470 '45/62'726 +
chr1 10470 10472 '48/65'738 +
...
```

Die ersten drei Spalten beinhalten Chromosom, Startposition und Endposition. Die vierte Spalte gibt Informationen über die Anzahl an methylierten Cytosinen (im Beispiel: 45 bzw. 48) und die genomische Abdeckung (*coverage*) (im Beispiel: 62 bzw. 65). Die nächste Spalte gibt den Grad der Methylierung an, welcher auch nach folgender Formel aus Spalte vier berechnet werden kann.

$$\frac{\text{AnzahlC(methyliert)}}{\text{coverage}} * 1000$$

Das + (-) steht für den Strang (*forward*-Strang: „+“; *reverse*-Strang: „-“), auf dem das CpG-Dinukleotid liegt.

Methylo 1 und Methylo 2 wurden am Institut für Humangenetik generiert und durch die Genominformatik auf einer internen Webseite des Universitätsklinikums Essens zur Verfügung gestellt (siehe Abschnitt 2.5.1). Methylierungsdaten einzelner CpG-Dinukleotiden und ganzer Regionen wie CGIs oder Genen konnten über diese Seite abgerufen und so in einer Datei gespeichert werden. Das Beispiel Methylo-Daten (2) zeigt wie diese Daten aufgebaut sind. Jede Zeile steht für ein CpG-Dinukleotid und besitzt vier Informationen: Chromosom, Position, Grad der Methylierung (0.0-1.0) und die Anzahl an Reads, die

in die Berechnung der Methylierung eingegangen sind, somit die genomische Abdeckung dieser genomischen Position.

Beispiel 2.7: Methylo-m-Daten (2)

```
<1> <2> <3> <4>
...
2 1653572 1.0 18
2 1653586 0.93 14
2 1653758 0.87 16
...
```

Die Oozyten-Methylome publiziert von Okae u. a. (2014) sind als tab-separierte Datei aufgebaut und beinhalten die folgenden fünf Spalten. Neben dem Chromosom, Start und End des CpG-Dinukleotids ist die Anzahl an Cytosinen und die Anzahl an Thyminen angegeben (siehe Beispiel Methylo-m-Daten (3)).

Beispiel 2.8: Methylo-m-Daten (3)

```
<1> <2> <3> <4> <5>
...
chr1 240656453 240656454 2 7
chr1 240656455 240656456 2 6
chr1 240656481 240656482 0 4
...
```

2.1.6 Chromatin-Daten

Bei den untersuchten Chromatin-Daten handelt es sich um nicht-veröffentlichte Daten innerhalb eines Forschungsverbundes, an dem das Institut für Humangenetik des Universitätsklinikums beteiligt ist. Hierbei wurden in Monozyten von zwei unterschiedlichen Spendern (interne IDs 43_Hm04_BIMo_Ct und 43_Hm05_BIMo_Ct) die folgenden Histonmodifikationen untersucht: H3K27ac, H3K4me1, H3K4me3, H3K36me3, H3K27me3 und H3K9me3. Zusätzlich wurde auch die Chromatinzugänglichkeit (*chromatin accessibility*, CA) untersucht. Diese Chromatin-Daten wurden mit Hilfe der Genominformatik auf einer internen Webseite des Universitätsklinikums Essens zur Verfügung gestellt und konnten mit Hilfe des IGV Browsers visualisiert werden.

2.1.7 Daten über repetitive Sequenzen

Daten über repetitive sequenzen (*repeats*) wurden über den UCSC Browser heruntergeladen. Innerhalb des UCSC Browsers wird hierzu das Programm RepeatMasker von Arian Smit verwendet (siehe Abschnitt A.1), was wiederum auf die Repbase Datenbank (Jurka,

2000) zugreift. Alle bekannte Klassen von repetitiven Sequenzen, wie z.B. SINEs (*short interspersed nuclear elements*), LINEs (*long interspersed nuclear elements*), LTRs (*long terminal repeat elements*), Mikrosatelliten und RNA-Repeats sind in diesem Datensatz vorhanden. Jede Zeile des Datensatzes steht für eine repetitive Sequenz. Das folgende Beispiel zeigt exemplarisch eine Zeile.

Beispiel 2.9: Repetitive Sequenzen

```
<1> <2> <3> <4> <5> <6> <7> <8> <9> <10><11><12><13><14> <15> <16> <17>
...
585 239 338 148 0 chr1 18906 19048 -249231573 + L2a LINE L2 2942 3104 -322 7
...
```

Jede Zeile unterteilt sich in 17 Spalten, wobei wiederum die erste Spalte nur zum Indizieren innerhalb des UCSC Browsers dient. Die folgenden vier Spalten beschreiben neben dem Score des Smith Waterman Alignments die Anzahl der Mismatches, der gelöschten Basen und der eingefügten Basen in Promille. Die Spalten sechs bis zehn geben genaue Informationen über die genomische Lage der repetitiven Sequenz. Neben dem Chromosom und der Start- bzw. Stopposition im Genom sind die Anzahl an Basen nach einem Treffer in einer genomischen Sequenz und der Strang angegeben. Die nächsten sechs Spalten geben weitere Informationen über die repetitive Sequenz an sich an, wie z.B. den Namen, die Klasse und die Familie. Zusätzlich werden noch Start- und Endposition der repetitiven Sequenz angegeben und die Anzahl an Basen nach einem Treffer. Die letzte Spalte beinhaltet eine ID, die ausschließlich innerhalb des Programms RepeatMasker benutzt wird. In dieser Arbeit werden nur die Informationen der folgenden Spalten benutzt: 6-8 und 10-13.

2.2 *In silico* Methoden

2.2.1 Nummerierung der CGIs

Die Benennung von CGIs innerhalb des UCSC Browsers erfolgt nach der Anzahl der CpG-Dinukleotide. Da es bei 27.537 humanen und 15.997 murinen CGIs vorkommen kann, dass mehrere CGIs auf Grund ihrer Anzahl von CpG-Dinukleotiden denselben Namen haben, wurden sowohl die humanen als auch die murinen CGIs fortlaufend von Chromosom 1 an eindeutig neu nummeriert. Um zwischen murinen und humanen CGIs zu unterscheiden, wurde neben der eindeutigen Zahl auch der Organismus in den Namen eines CGIs mit aufgenommen. 123_1_hg19 und 23_1_mm10 sind zum Beispiel Namen für CGIs, wie sie in dieser Arbeit verwendet werden. Die zweite Zahl gibt an, ob ein CGI

mit mehreren Genen bzw. proteinkodierenden Regionen überlappt, was jedoch erst im nächsten Abschnitt (2.2.2) berechnet wird. Durch diese zusätzliche Information ist es möglich, dass keine Informationen, die in 2.2.2 berechnet werden, verloren gehen. Das folgende Beispiel zeigt die überarbeitete Datei mit drei zusätzlichen Spalten: CGI ID (Spalte eins), Organismus (Spalte zwei) und die Datenquelle (Spalte 13).

Beispiel 2.10: Neue CGI-Datei

```

<1>      <2>  <3>  <4>      <5>  <6>      <7>  <8> <9> <10> <11> <12> <13>
...
1_1_hg19 hg19 chr1 28735 29810 CpG:116 1075 116 787 21.6 73.2 0.83 UCSC
2_1_hg19 hg19 chr1 135124 135563 CpG:30 439 30 295 13.7 67.2 0.64 UCSC

```

2.2.2 Bestimmung der relativen Lage von CGIs zu proteinkodierenden Regionen (CCDS)

Zur Bestimmung der relativen Lage eines CGIs in Relation zu einem Gen, wurden fünf Klassen eingeführt, die diese Lage beschreiben (graphische Darstellung siehe Abb. 2.1).

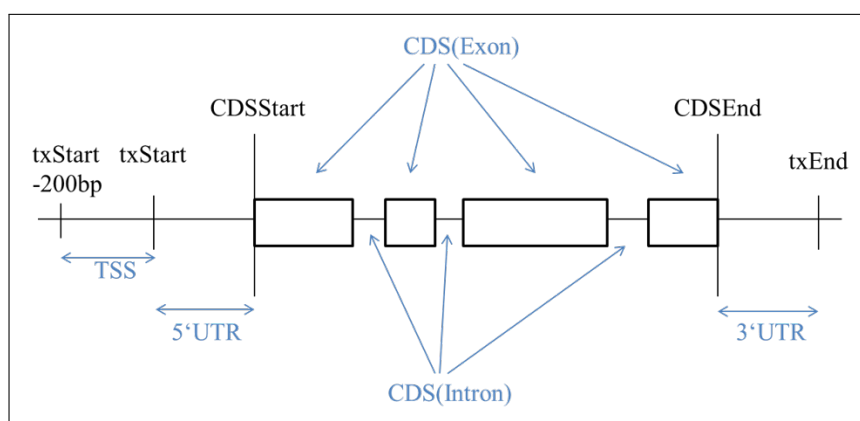


Abbildung 2.1: Schematische Darstellung von möglichen Positionen eines CGIs in Abhängigkeit von einem CCDS

schwarz: schematische Darstellung einer CCDS mit Transkriptionsstart (txStart), Start des CDS (CDSstart), Ende des CDS (CDSend) und Transkriptionsende (txEnd); blau: mögliche Klassen, in denen ein CGI liegen kann: TSS (Region 200bp *upstream* vom Transkriptionsstart), 5'UTR, CDS(Exon), CDS(Intron) und 3'UTR (bereits veröffentlicht in Rademacher u. a. (2014)).

Die Klasse TSS (Transkriptionsstartstelle) beschreibt die Region 200bp *upstream* vom Transkriptionsstart. Als 5'UTR (untranslatierte Region) wird die Region zwischen Transkriptionsstart und CDS-Start beschrieben. Liegt ein CGI innerhalb eines Exons, so gehört es zu der Klasse CDS(Exon). Liegt es innerhalb eines Introns, gehört es zu der Klasse

CDS(Intron). Die Region zwischen CDS-Ende und Transkriptionsende wird 3'UTR genannt. Ein CGI kann entweder zu keiner Klasse, zu einer Klasse oder zu mehreren Klassen gehören. Insgesamt wurden durch diese Methode zwei unterschiedliche Dateien als Ausgabe generiert. Das folgende Beispiel zeigt die „Lokalisierungsdatei“. Neben der eindeutigen ID des CGIs (Spalte 1), ist in dieser Datei auch die Klasseneinteilung eines CGIs beschrieben. Wenn ein CGI zu einer gewissen Klasse gehört, wird in der entsprechenden Spalte ein „yes“ eingetragen, sollte dies nicht der Fall sein, ein „no“ (Perl-Skripte siehe beigelegte Daten-CD).

Beispiel 2.11: Lokalisierungsdatei

```
<1>      <2><3><4><5><6><7>  <8>  <9>  <10><11> <12>
...
19_1_hg19 no no yes yes no no 8017 8263 - hg19 N0C2L
20_1_hg19 yes yes yes yes no no -8035 306 - hg19 N0C2L
20_2_hg19 yes yes yes yes yes yes 1760 -6581 + hg19 KLHL17
...
```

Die Klassen sind in der folgenden Reihenfolge in der Datei codiert: TSS, 5'UTR, CDS(Exon), CDS(Intron) und 3'UTR. In der siebten Spalte steht noch zusätzlich die Information, ob das CGI mit der gesamten CDS überlappt. Die Spalten acht und neun geben die Distanz des CGIs zum CDS Start und zum CDS Ende an (siehe Abbildung 2.2). Die letzte Spalte gibt das Gen an, mit dem das CGI überlappt. Falls keine Überlappung mit einem Gen vorliegt, wird in die Spalte „NONE“ eingetragen. Bei der zweiten Datei handelt es sich um die „Nachbarschaftsdatei“ (siehe folgendes Beispiel).

Beispiel 2.12: Nachbarschaftsdatei

```
<1>      <2>  <3>  <4>  <5>  <6>  <7>
...
19_1_hg19 25236 SAMD11 N0C2L KLHL17 9364 hg19
20_1_hg19 33193 SAMD11 N0C2L KLHL17 -6688 hg19
20_2_hg19 -366 N0C2L KLHL17 PLEKHN1 -778 hg19
...
```

Wie dem Namen schon zu entnehmen ist, beinhaltet diese Datei Informationen über die Umgebung eines CGI. In der ersten Spalte steht wiederum die eindeutige ID des CGIs. Die Spalten zwei und sechs beinhalten Distanzen, wobei in Spalte zwei die Distanz vom Start des CGIs bis zum Transkriptionsstart des vorherigen Gens (*upstream*) und in Spalte sechs die Distanz vom Ende des CGIs bis zum Transkriptionsstart des nächsten Gens (*downstream*) codiert ist (schematische Darstellung in Abbildung 2.2 (B)).

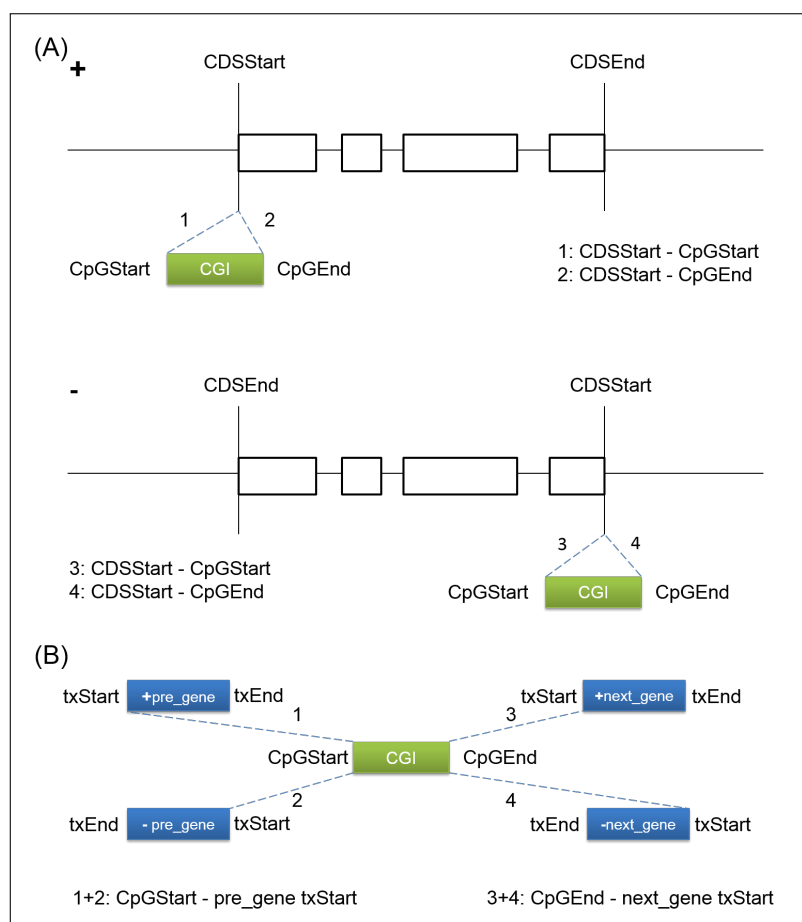


Abbildung 2.2: Schematische Darstellung und Berechnung der Distanzen von Lokalisierungs- und Nachbarschaftsdatei

Die Abbildung stellt eine graphische Darstellung der berechneten Distanzen, inklusive ihrer Berechnung, dar. (A) zeigt die Distanzberechnung, wenn ein CGI (grün) innerhalb einer CCDS liegt (schwarze Darstellung mit CDS Start/Ende und Exon-Intron-Struktur). Hierbei wird unterschieden, auf welchem Strang das Gen liegt (+/-). In (B) werden die Distanzen eines CGIs zu benachbarten Genen dargestellt. Auch hier werden die Distanzen vom CGI aus berechnet, vom CGI Start aus *upstream* zum Transkriptionsstart des vorherigen Gens (pre_gene; blau) und vom CGI Ende aus *downstream* zum Transkriptionsstart des nächsten Gens (next_gene; blau).

Die Gennamen der umliegenden Gene und des Gens, mit dem das CGI überlappt, sind in den Spalten drei bis fünf eingetragen, wobei Spalte drei den Namen des vorherigen Gens (pre_gene; *upstream*) und Spalte fünf des nächsten Gens (next_gene; *downstream*) beinhaltet. Zwischen diesen beiden Genen steht in Spalte vier der Name des Gens, in dem das CGI liegt. Sollte die Position eines CGIs nicht mit der Position einer CCDS überlappen, so wird anstatt eines Gennamens „NONE“ in die vierte Spalte eingetragen. Dies kann auch in den Spalten drei und fünf der Fall sein, falls ein CGI nah am Telomer eines Chromosoms

liegt. Sollte es wie bereits in Abschnitt (2.2.1) erläutert der Fall sein, dass ein CGI mit mehreren Genen überlappt, so wird in die Datei jeweils eine neue Zeile eingefügt und die zweite Zahl der eindeutigen ID hochgezählt. Bei den CGIs 123_1_hg19 und 123_2_hg19 handelt es sich somit um dasselbe CGI, was jedoch mit unterschiedlichen Genen bzw. proteinkodierenden Regionen überlappt (Perl-Skripte siehe beigelegte Daten-CD).

2.2.3 Sequenzvergleich

Ein Sequenzvergleich zwischen humanen und murinen CGIs, die im selben CCDS liegen, wurde mit Hilfe des Programmes *blast2seq* (*blast two sequences*) durchgeführt (Zhang u. a., 2000). Um festzustellen, ob ein CGI Sequenzähnlichkeiten zu anderen Regionen im Genom hat, wurde das *basic local alignment search tool* (BLAST) verwendet. In dieser Arbeit wurde diese Suche als diskontinuierliche MEGA-BLAST-Suche mit Standardeinstellungen gegen alle assemblierten *scaffolds* durchgeführt (Database Name: Genome[all assemblies scaffolds]; Description: Homo sapiens all assemblies [GCF_00000140522 GCF_000306695.1 GCF_000002135.2 GCF_000002125.1] scaffolds in NCBI Annotation Release 104; Program: BLASTN 2.2.28). Diese Analyse wird auch zur Vorhersage des Ursprungs eines CGIs benutzt (siehe Abschnitt 2.2.4, Perl-Skripte siehe beigelegte Daten-CD).

2.2.4 Evolutionsanalyse

Es ist bekannt, dass CGIs im Laufe der Evolution entstehen können. Ein Beispiel hierfür ist das CpG85 des humanen *RB1*-Gens (Kamber u. a., 2013). Das BLAT-Tool (Kent, 2002) auf der Seite des UCSC Browsers wurde benutzt, um herauszufinden, in welchen Primaten bestimmte CGIs vorhanden sind. Die BLAT-Suche wurde für insgesamt sieben unterschiedliche Primaten durchgeführt: Schimpanse (CSAC 2.1.4/panTro4), Gorilla (gorGor3.1/gorGor3), Orang-Utan (WUGSC 2.0.2/ponAbe2), Gibbon (GGSC Nleu3.0/nomLeu3), Rhesusaffe (BGI CR_1.0/rheMac3), Marmosette (WUGSC3.2/calJac3) und Galago (Broad/otoGar3). Für die Suche wurden die Sequenzen der CGIs und auch Sequenzen von flankierenden Exons benutzt.

2.2.5 Finden von Retrokopie-assoziierten CGIs

Neue CGIs können durch Retrotransposition entstehen. Um dieses biologische Phänomen zu untersuchen, wurden die Datensätze der CGIs und der Retrokopien miteinander verglichen. Hierbei wurde nach Überlappungen der genomischen Positionen gesucht. Falls eine Überlappung vorliegt, wurde eine neue Datei generiert, die aus den CGI-Daten und den Retrokopie-Daten besteht. Hierzu wurde einfach die Zeile des Retrogen-Eintrags (2.1.3) an die Zeile des überlappenden CGIs (2.2.1) angehängt (Perl-Skripte siehe beigefügte Daten-CD).

2.2.6 Analyse der Umgebung von Retrokopien in Hinblick auf repetitive Sequenzen

Repetitive Sequenzen können die Methylierung einer Region beeinflussen. Um einen möglichen Einfluss festzustellen bzw. auszuschließen, wurden die Datensätze der Retrokopien (die Regionen 1000bp *upstream* und 1000bp *downstream* einer Retrokopie) und der repetitiven Sequenzen miteinander verglichen. Hierbei wurde ebenfalls nach Überlappungen in den genomischen Positionen gesucht. Falls eine Überlappung vorliegt, wird der Zeile der Region Informationen zu der jeweiligen repetitiven Sequenz angehängt (Perl-Skripte siehe beigefügte Daten-CD).

2.2.7 Berechnung der Methylierung von bestimmten Regionen

Die Methylierung bestimmter Regionen wurde mit Hilfe der Methylom-Daten aus Abschnitt 2.1.5 und der Programmiersprache Perl berechnet (Perl-Skripte siehe beigefügte Daten-CD). Es wurde jeweils immer die durchschnittliche Methylierung einer bestimmten Region nach folgender Formel berechnet:

$$Methylierung_Region = \frac{1}{n} \sum_{i=1}^n Methylierung_CpG_i$$

Um die Methylierung der Umgebung einer Retrokopie zu berechnen, wurde der Mittelwert der Methylierung von 1000bp *downstream* und *upstream* der Retrokopie berechnet.

2.2.8 Statistik

Der Korrelationskoeffizient nach Pearson wurde mit Hilfe der Statistik-Software R und dem folgenden Aufruf berechnet:

```
cor(data1, data2, method = "pearson").
```

Die Clusteranalyse k-means wurde, nach dem Algorithmus von Hartigan u. Wong (1979), ebenfalls mit der Statistik-Software R durchgeführt. Hierbei wurde der folgende Aufruf benutzt:

```
fit <- kmeans(data, 4, algorithm = „Hartigan – Wong“)  
aggregate(data, by = list(fit$cluster), FUN = mean)  
mydata <- data.frame(data, fit$cluster).
```

2.3 Material

2.3.1 Chemikalien

Die Chemikalien wurden, soweit nicht anders vermerkt, von den Firmen Invitrogen (Karlsruhe), Merck (Darmstadt), Peqlab (Erlangen) und Roche (Mannheim) in *pro analysis* Qualität bezogen.

2.3.2 Standardlösungen

TAE-Puffer: 40mM Tris-Acetat
 1mM EDTA
 pH 8,0

2.3.3 Größenstandards

DNA-Größenstandards

MassRuler™ Low Range DNA Ladder	Thermo Fisher Scientific
pUC19 DNA/ <i>MspI</i> Marker	Thermo Fisher Scientific

GeneScan

GeneScan-LIZ 120 Size Standard	Applied Biosystems
--------------------------------	--------------------

2.3.4 Enzyme

<i>AmpliTaq</i> Gold Polymerase	Roche (Mannheim)
HotStarTaq Master Mix Kit	Qiagen (Hilden)
GoTaq Green Master Mix	Promega (Mannheim)
DNAse RQ1	Promega (Mannheim)
Reverse Transcriptase	Applied Biosystems
RNAse Inhibitor	Applied Biosystems

2.3.5 Oligonukleotide

Mit Hilfe folgender Programme (URL siehe A.1) wurden die verwendeten Primer gesucht:

Primer3
MethPrimer

Die Sequenzen der Primer sowie die Annealing Temperaturen und die Fragmentgrößen sind im Anhang (A.2) aufgelistet.

Folgende Programme (URL siehe A.1) zum Durchführen von *in silico* PCRs wurden verwendet:

UCSC *in silico* PCR
BiSearch

Alle Oligonukleotide wurden zusätzlich noch auf ihre Spezifität überprüft, indem mit ihren Sequenzen eine BLAT-Suche (URL siehe A.1) durchgeführt wurde.

2.3.6 Humane Proben

Für alle humanen Proben lag eine Einverständniserklärung in Essen oder bei einem Kollaborationspartner vor, sowie ein positives Ethikvotum.

2.3.6.1 Monozyten

Humane Monozyten wurden aus Blutspenden von insgesamt 24 gesunden männlichen Spendern isoliert und anonymisiert (Instituts IDs M55900, 43_Hm01_BIMo_Ct, 43_Hm04_BIMo_Ct, 43_Hm05_BIMo_Ct, R1-R17, P1-P3, K1 und K2). Die Probenentnahme und Monozyten-Isolierung wurde durch Herrn Dr. Stefan Wallner am Institut für Klinische Chemie und Laboratoriumsmedizin am Universitätsklinikum Regensburg durchgeführt.

2.3.6.2 Trios

In dieser Arbeit wurde DNA von Trios (Mutter, Vater, Kind) und RNA (vom Kind, wenn vorhanden) untersucht. Hierbei handelt es sich um Trios einer NGS-Studie (*next generation sequencing*) von Prof. Dr. Dagmar Wieczorek.

2.3.6.3 Patienten mit einem Multilocus-*Imprinting*-Defekt (MLID)

Es wurden sechs Patienten mit einem Multilocus-*Imprinting*-Defekt (MLID) innerhalb dieser Arbeit untersucht. Hierbei handelt es sich um Patienten, die innerhalb des BMBF-Konsortiums *Diseases caused by imprinting defects* (01GM1513) rekrutiert worden sind. Einen Silver-Russel-Phänotyp weisen drei der sechs Patienten auf (MLID1, MLID3 und MLID5), zwei Patienten zeigen einen Beckwith-Wiedemann-Phänotyp (MLID2 und MLID6) und ein Patient zeigt einen Pseudohypoparathyreoidismus Typ 1B-Phänotyp.

2.3.6.4 Gewebe

Gewebeproben (Plazenta) wurden von Frau PD Dr. Alexandra Gellhaus (Klinik für Frauenheilkunde und Geburtshilfe, Universitätsklinikum Essen) für Forschungszwecke zur Verfügung gestellt. Hierbei handelt es sich um Material von Aborten bzw. natürlichen Geburten, was für die Forschung benutzt werden darf. Insgesamt wurden 15 Gewebeproben (jeweils fünf Proben aus den unterschiedlichen Trimenons) und dazugehörige RNA-Proben bereitgestellt. Eine mögliche mütterliche Kontamination der DNA wurde mithilfe des PowerPlex® 16 Systems (Promega) ausgeschlossen (durchgeführt durch Frau Melanie Heitmann). Zusätzlich zu diesen Proben wurden noch DNA und RNA aus Plazenta von der Baby Bio Bank des University College London bestellt. Um einen mögliche Ursprung klären zu können, wurde auch DNA aus Blut von den Eltern bestellt.

2.4 *In vitro* Methoden

2.4.1 *Whole-Genome Bisulfite Sequencing* (WGBS)

Das *Whole-Genome Bisulfite Sequencing* wurde von Dr. Michael Zeschnigk am Institut für Humangenetik des Universitätsklinikums Essen in Zusammenarbeit mit dem BioChip-Labor etabliert. Die Methode wurde in Rademacher u. a. (2014) ausführlich beschrieben und lässt sich in den folgenden Schritten zusammenfassen:

- *library preparation*
- Sequenzierung (Illumina HiSeq2500)
- Konvertierung der Daten

2.4.2 Arbeiten mit DNA

2.4.2.1 Isolierung von DNA

DNA-Isolierung aus Mundschleimhaut und Blut

DNA-Isolierung aus Mundschleimhaut und Blut wurde mit Hilfe des BioRobot EZ1 (Qiagen) durchgeführt. Die Isolierung erfolgte hierbei vollautomatisiert unter Verwendung

von Magnet-Partikeln. In folgende drei Schritte lässt sich die Isolierung mit dem Roboter unterteilen:

- Lyse der Proben
- Bindung der DNA an magnetische Partikel
- Waschen und Elution der DNA

Bei der Isolierung wurde nach Herstellerangaben verfahren.

DNA-Isolierung aus Gewebe

Die Isolierung von DNA aus Gewebe (Plazenta) wurde mit Hilfe der FlexiGene-Methode (Qiagen) durchgeführt. Hierbei wurde nach dem folgenden Protokoll vorgegangen:

- Thermomixer auf 56°C vorheizen
- Für jede Probe zwei 1,5ml Schraubdeckel-Gefäße und ein 1,5ml Schnappdeckel-Gefäß vorbereiten (beschriften)
- Mastermix ansetzen: je Probe 500µl FG2-Puffer und 20µl Proteinase K
- Gewebestück (20mg) mit 1000µl FG1-Puffer waschen und 5min auf den Roller legen, anschließend kurz abzentrifugieren und den Puffer abpipettieren
- 250µl Mastermix zum Gewebestück hinzufügen und zusammen in den Mörser überführen
- Gewebestück mit Mörser und Stößel homogenisieren
- Homogenisat mit einer Glaspipette in ein Schnappdeckel-Tube überführen
- Mörser und Stößel mit 250µl Mastermix spülen und ebenfalls in Schnappdeckel-Gefäß überführen
- 10µl Proteinase K hinzupipettieren und die Probe sofort im Thermomixer für mindestens 4h bei 56°C inkubieren (850rpm)
- DNA mit 510µl Isopropanol fällen, danach mit der Glaspipette „angeln“ und anschließend zum Waschen in ein 0,5ml Schnappdeckel-Tube mit 300µl 70%igem Ethanol eingetaucht. Zum Trocknen der DNA Glaspipette umgedreht in einen Ständer platzieren (Alternative: bei sehr geringer DNA Menge, die Probe für 2min bei 2000g zentrifugieren. Überstand verwerfen und das Pellet mit 500µl 70%igem Ethanol waschen. Überstand wiederum verwerfen und erneut 500µl Ethanol hinzugeben und für 2min bei 2000g zentrifugieren. Überstand verwerfen und das DNA Pellet ca. 10min trocknen lassen)
- Das getrocknete Pellet wird in 30-100µl FG3 Puffer aufgenommen
- Lösen der DNA für 1h bei 65°C unter Schütteln (850rpm)

- DNA wird über Nacht auf dem Roller (Raumtemperatur) weiter gelöst und anschließend Ausbeute und Reinheit bestimmt (2.4.2.2)

2.4.2.2 Konzentrationsbestimmung von DNA

Die Konzentration von DNA wurde photometrisch bestimmt. Hierzu wurde ein NanoDrop ND-1000 (NanoDrop Technologies) benutzt, welcher die Absorption bei einer Wellenlänge von 260nm und 280nm misst. Eine optische Dichte (OD) von 1 bei einer Wellenlänge von 260nm (OD_{260}) entspricht einer DNA-Konzentration von 50 μ g/ml. Der Quotient OD_{260}/OD_{280} gibt den Grad der Reinheit an. Von reiner DNA spricht man bei einem Wert von 1,8.

2.4.2.3 Polymerasekettenreaktion (PCR)

Mittels der Polymerasekettenreaktion nach Mullis u. a. (1986) wurden bestimmte Sequenzabschnitte von genomischer und Bisulfit-behandelter DNA amplifiziert. Standardmäßig wurde die GoTaq[®] G2 DNA Polymerase (Promega) benutzt. Ein Ansatz von 25 μ l Endvolumen enthielt 100ng genomische DNA, jeweils 0,4 μ M Primer (for- und rev-Primer), jeweils 200 μ M dNTPs (dATP, dCTP, dTTP und dGTP), 1,5mM Magnesiumchlorid ($MgCl_2$), 1x GoTaq Green Reaktionspuffer und 1,25U GoTaq[®] G2 DNA Polymerase. Die eigentliche PCR-Reaktion erfolgte im Regelfall wie folgt im Thermocycler:

2min	95°C	(Polymeraseaktivierung)	} 35 Zyklen
30sek	95°C	(Denaturierung)	
30sek	50-60°C	(Annealing)	
45sek	72°C	(Elongation)	
5min	72°C	(finale Elongation)	
∞	4°C		

Für GC-reiche Sequenzen, bei denen die oben genannte PCR nicht funktionierte, wurde eine PCR mit Betain und 7-deaza-dGTP-dNTP etabliert (Rees u. a., 1993). Durch das Betain wird der Schmelztemperatur-Unterschied zwischen AT und GC angeglichen und das 7-deaza-dGTP-dNTP reduziert nach Einbau die Anzahl an Wasserstoffbrückenbindungen zur Base

Cytosin. Der Reaktionsansatz für eine Betain-PCR von 25µl Endvolumen enthielt 130ng genomische DNA, jeweils 0,4µM Primer (for- und rev-Primer), jeweils 80µM dNTPs (dATP, dCTP und dTTP), 32µM dGTP, 48µM 60% 7-deaza-2'-deoxy-guanosine-5'-triphosphate (Roche), 1,5mM MgCl₂, 0,5M Betain (USB Corporation), 1x Green GoTaq Reaktionspuffer und 5U GoTaq G2 DNA Polymerase. Das PCR-Programm unterteilt sich in die folgenden Schritte und erfolgte im Thermocycler:

2min	95°C	(Polymeraseaktivierung)	
30sek	96°C	(Denaturierung)	} 45 Zyklen
30sek	50-60°C	(Annealing)	
45sek	72°C	(Elongation)	
7min	72°C	(finale Elongation)	
∞	4°C		

Bisulfit-behandelte DNA wurde mit dem HotStarTaq Master Mix (Qiagen) amplifiziert. Ein Ansatz mit 25µl Endvolumen setzte sich aus 1µl Bisulfit-behandelter DNA, 1x HotStarTaq Master Mix und jeweils 0,2µM Primer (for- und rev-Primer). Die PCR-Reaktion erfolgte im Thermocycler in der Regel wie folgt:

15min	95°C	(Polymeraseaktivierung)	
30sek	95°C	(Denaturierung)	} 40 Zyklen
30sek	50-60°C	(Annealing)	
45sek	72°C	(Elongation)	
10min	72°C	(finale Elongation)	
∞	4°C		

Die zweite PCR (RePCR) folgt auf die PCR von Bisulfit-behandelter DNA. Es wurde ebenfalls der HotStarTaq Master Mix (Qiagen) zur Amplifikation verwendet. Ein Ansatz mit 50µl Endvolumen setzte sich aus 3µl PCR-Produkt, 1x HotStarTaq Master Mix und 0,2µM Primer (for- und rev-Primer) und wurde durch das folgende PCR-Programm (*twostep* PCR) im Thermocycler amplifiziert:

15min	95°C	(Polymeraseaktivierung)	} 40 Zyklen
30sek	95°C	(Denaturierung)	
45sek	72°C	(Elongation)	
10min	72°C	(finale Elongation)	
∞	4°C		

Im Anhang (siehe A.2) sind alle benutzten Primer inklusive der Primersequenzen, Annealingtemperaturen, Zyklusanzahlen und eventuelle Zugaben von Betain aufgelistet. Auch PCR-Programme, die von den soeben beschriebenen Standard-PCRs abweichen, sind im Anhang zu finden (siehe A.6).

2.4.2.4 Auftrennung von DNA in Agarosegelen

DNA-Fragmente wurden mittels Gelelektrophorese aufgetrennt. Hierzu wurden 1-2%ige Agarose-Minigele mit 1 x TAE-Puffer und 0,4µg/ml Ethidiumbromid hergestellt. Bei 100-120 Volt erfolgte die Auftrennung der Fragmente, wobei neben den Proben Größenstandards (2.3.3) zur Ermittlung der spezifischen Fragmentgröße mit aufgetragen wurden.

2.4.2.5 Aufreinigung von PCR-Produkten

Zur Aufreinigung von PCR-Produkten wurden die folgenden Techniken benutzt:

Millipore Aufreinigung

Das Millipore Multiscreen Filtration System wurde benutzt und die Aufreinigung der PCR-Produkte nach Anleitung des Herstellers durchgeführt.

ExoSAP-IT

PCR-Produkte wurden mittels USB ExoSAP-IT® PCR Product Cleanup (Affymetrix) aufgereinigt. Hierzu wurden 5µl der PCR-Produkte mit 2µl ExoSAP-IT® versetzt. Die Aufreinigung erfolgte in einem Thermocycler, indem die PCR-Produkte für 15min bei 37°C inkubiert wurden und anschließend die Reaktion bei 80°C (ebenfalls 15min) terminiert wurde.

Elution von PCR-Produkten aus Agarosegelen

Aus dem Agarosegel ausgeschnittene PCR-Produkte wurden anschließend mit dem Wizard®

SV Gel and PCR Clean-Up System von Promega aufgereinigt. Dies erfolgte nach Protokoll des Herstellers.

2.4.2.6 Sequenzierung nach Sanger

Die Sequenzierung von PCR-Produkten erfolgte mittels Kettenabbruchverfahren nach Sanger u. a. (1977) und wurde mit dem BigDye[®] Terminator v1.1 Cycle Sequencing Kit (Life Technologies) durchgeführt. Ein Ansatz hatte ein Gesamtvolumen von 10µl und setzte sich aus 1µl (Konzentration 5µM) Sequenzierprimer (for- oder rev-Primer), 4µl BigDye[®], 1-3µl aufgereinigte PCR-Produkte (abhängig von der Bande im Gel und der Aufreinigungsmethode) und 2-4µl H₂O. Die Sequenzierreaktion erfolgte durch folgende Schritte im Thermocycler:

1min	96°C		
10sek	96°C	(Denaturierung)	} 25 Zyklen
5sek	50-60°C	(Annealing)	
4min	60°C	(Elongation)	
∞	4°C		

Die Produkte der Sequenzierreaktion wurden mit 10µl H₂O aufgefüllt und mit Hilfe von Sephadex Gelfiltrations Säulen (GE Healthcare) aufgereinigt. Anschließend wurde das Eluat (20µl) mit 15µl Formamid versetzt und auf dem ABI 3130xl Genetic Analyzer (Life Technologies) sequenziert. Die Auswertung der Daten erfolgte mit den Programmen Sequencing Analysis (Applied Biosystems) und Geneious (Biomatters).

2.4.2.7 Bisulfitkonvertierung von DNA

Um die Methylierung von DNA genauer zu untersuchen, wurde die DNA mit Bisulfit behandelt, hierbei werden unmethylierte Cytosine deaminiert und zu Uracilen konvertiert. Bei einer anschließenden PCR der Bisulfit-behandelten DNA, wird Uracil durch Thymin ersetzt. Methylierte Cytosine im CpG Kontext sind gegen diese chemische Umwandlung durch ihre Methylgruppe geschützt. Methylierte und unmethylierte Cytosine können somit bei einer nachfolgenden Sequenzierung (z.B. auf dem Roche/454 Genome sequencer (GS) Junior) unterschieden werden.

Für die Bisulfitkonvertierung wurde das EZ DNA Methylation-Gold™ Kit (Zymo Research) benutzt, wobei nach Herstellerangaben verfahren wurde. Jeweils 500ng DNA wurden in Bisulfit-behandelte DNA konvertiert.

2.4.2.8 Methylierungsanalyse von Bisulfit-behandelter DNA

Zur Methylierungsanalyse wurde der Roche/454 Genome sequencer (GS) Junior benutzt. Die hochquantitative Hochdurchsatz-Bisulfitsequenzierung (*deep bisulfite sequencing*) wurde am Essener Institut für Humangenetik etabliert und in Beygo u. a. (2013) beschrieben. Zusätzlich hat Roche in Zusammenarbeit mit dem Institut für Humangenetik eine *Application Note* veröffentlicht, in der ausschließlich die Methylierungsanalyse mittels Roche 454 beschrieben wird (Horsthemke u. a., 2013). Sie unterteilt sich in folgende drei Schritte:

- Herstellung von *Amplicon libraries*
- Aufreinigung der *Amplicon libraries* und Emulsions-PCR (emPCR)
- Sequenzierlauf auf dem Roche/454 GS Junior

Die Herstellung von *Amplicon libraries* erfolgt durch zwei aufeinanderfolgende PCRs (siehe Abbildung 2.3 und Abschnitt 2.4.2.3. Die erste PCR erfolgt mit spezifischen Primern, die mit universellen *tag*-Sequenzen an ihrem 5'Ende modifiziert wurden. Die genauen Primersequenzen inklusive der Annealingtemperaturen und der Fragmentgrößen sind im Anhang A.2.2 zu finden. Die PCR-Fragmente wurden mittels Gelelektrophorese (2.4.2.4) überprüft und anschließend aus dem Gel eluiert (2.4.2.5). Bei der anschließenden PCR, einer so genannten RePCR, wurden dann Barcode-Sequenzen (MIDs, *multiplex identifier*; siehe Anhang A.2.5), eine 4bp *Key*-Sequenz und A- bzw. B-Primer angehängt. Auch diese Produkte wurden mittels einer Gelelektrophorese überprüft. Die *Key*-Sequenz wird später für Filterschritte innerhalb der Datenanalyse benötigt, die Primer für die klonale Amplifikation innerhalb der Emulsions-PCR (emPCR). Das PCR-Produkt wurde mit H₂O auf ein Gesamtvolumen von 45µl aufgefüllt. Die weitere Probenvorbereitung und der Sequenzierlauf auf dem Roche/454 GS Junior wurden durch Frau Melanie Heitmann, Frau Sabine Kaya oder Frau Claudia Mertel (Institut für Humangenetik, Universitätsklinikum Essen) durchgeführt. Während der Sequenzierung auf der Roche/454 GS Junior Plattform werden nacheinander Lösungen mit den vier fluoreszenzmarkierten Basen über die einzelnen *wells*, die jeweils mit einem *bead* belegt sind, gespült. Wenn eine Base eingebaut wird, entsteht ein Signal, dass mit Hilfe einer speziellen Kamera detektiert wird. Dies geschieht

innerhalb eines Sequenzierlaufes für jedes *well*, das anschließend über so genannte *flowgrams* in die entsprechenden Sequenzen umgerechnet wird.

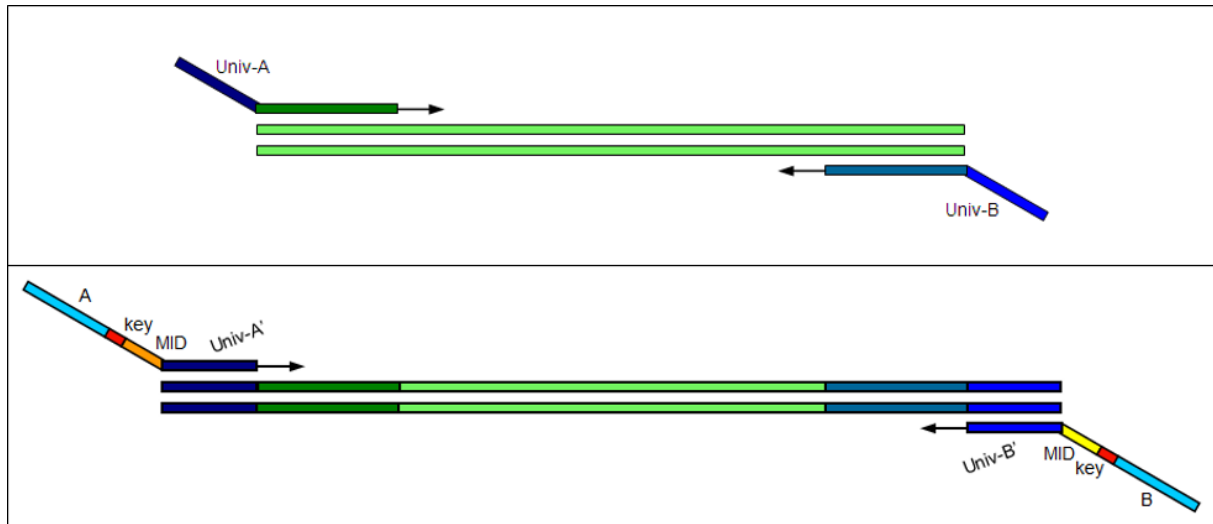


Abbildung 2.3: Schematische Darstellung der zwei aufeinanderfolgenden PCRs bei der Herstellung von *Amplicon libraries*. Die Graphik wurde der Roche *Sequencing Application Note* entnommen (Horsthemke u. a., 2013).

Im oberen Teil der Graphik ist die erste PCR bei der Herstellung von *Amplicon libraries* dargestellt, die Ampliconsequenz in hellgrün, die spezifischen Primer mit den universellen *tag*-Sequenzen oberhalb der Ampliconsequenz in dunkelgrün (spezifische Primer) und dunkelblau (universelle *tag*-Sequenzen). Der untere Teil zeigt die zweite PCR. Hierbei werden Barcode-Sequenzen (MIDs, *multiplex identifier*; dargestellt in gelb und orange), eine *Key*-Sequenz (rot) und die Primer A und B angehängt (hellblau)

2.4.3 Arbeiten mit RNA

2.4.3.1 Isolierung von RNA

RNA-Isolierung mit dem PAXgene® Blood RNA Kit (Blut)

RNA aus Blut, dass in PAXgene Blood RNA *tubes* entnommen wurde, wurde mit Hilfe des PAXgene® Blood RNA-Systems (PreAnalytiX) nach Herstellerangaben des Handbuchs vom April 2008 aufgereinigt und die RNA isoliert. Dieses System stabilisiert intrazelluläre RNA und ist somit laut Hersteller für RT-PCRs gut geeignet.

RNA-Isolierung mit TRIzol® (Blut)

RNA aus EDTA-Blut wurde mit Hilfe von TRIzol® (Life Technologies) isoliert. Die Isolierung wurde nach Herstellerangaben durchgeführt.

2.4.3.2 Konzentrationsbestimmung von RNA

Wie auch bei DNA (2.4.3.2), wird die Konzentration von RNA photometrisch mit dem NanoDrop ND-1000 (NanoDrop Technologies) gemessen. Eine optische Dichte von 1 bei einer Wellenlänge von 260nm entspricht einer RNA-Konzentration von 40µg/ml. Ebenfalls wird der Grad der Reinheit von RNA mit dem Quotienten OD_{260}/OD_{280} bestimmt. Dieser liegt bei reiner RNA zwischen 1,8 und 2,1.

2.4.3.3 DNase-Behandlung

Zur DNase-Behandlung wurde ein RQ1-Verdau mit Hilfe des RQ1 Kit (Promega) durchgeführt. RNA (2000ng) wurde in einem 20µl Ansatz mit 2µl RQ1-Reaktionspuffer und 2µl DNase RQ1 in einem Thermocycler bei 37°C für 30min inkubiert. Nach Zugabe von 2µl *stop solution* wurde die Reaktion bei 65°C für 10min terminiert.

2.4.3.4 Reverse Transkription (RT)

Nach dem RQ1-Verdau wurde die RNA mittels reverser Transkription in cDNA umgeschrieben. Für den *RPS2P32*-Locus wurden für jeden Stang eine strangspezifische reverse Transkriptase mit Hilfe der SuperScript™ III reversen Transcriptase (Invitrogen) nach Herstellerangaben durchgeführt. Für jede informative Probe wurden 2µg RNA (PAX-Blut) mit einem strangspezifischen Primer 9261_1_hg19_RT_F bzw. 9261_1_hg19_RT_R (Primersequenz siehe Anhang A.2.3) revers transkribiert.

Für den *IGF2BP3*-Locus erfolgte die reverse Transkription mit dem GeneAmp® RNA PCR Kit von Applied Biosystems (Life Technologies). Für jede informative Probe wurden 2µg RNA (PAX-Blut oder Plazenta) mit *random hexamers* reverse transkribiert (50µl Ansatz) wie in Kanber u. a. (2009a) beschrieben.

Um Kontaminationen sowohl von der RNA als auch von den Reagenzien auszuschließen, wurden für jeden Ansatz zwei Kontrollen mitgeführt: ein Ansatz mit RNA aber ohne Reverse Transkriptase und ein Ansatz mit Reverser Transkriptase aber ohne RNA.

2.4.3.5 Amplifikation der cDNA (RT-PCR)

Im Anschluß an die reverse Transkription wurde mit spezifischen Primern eine PCR an der cDNA durchgeführt. Als zusätzliche Kontrolle, neben den bereits unter 2.4.3.4 beschriebenen Kontrollen, wurde bei der Expressionsanalyse eine intronüberspannende Region des β -Aktin Gens mit spezifischen Primern amplifiziert, falls es nicht möglich war intronüberspannende Primer für das zu untersuchende Gen zu etablieren. Mit Hilfe der Expressionsanalyse des β -Aktin Gens kann eine mögliche DNA-Kontamination ausgeschlossen werden. Das RT-PCR-Produkt des β -Aktin Gens hat eine Größe von 496bp, wohingegen das PCR-Produkt bei DNA ca. 100bp größer ist (591bp) (Buiting u. a., 2001).

Die 50µl Ansätze wurden parallel angesetzt, wobei jeweils zwischen 12µl und 16µl des RT-Ansatzes (240-320ng cDNA) eingesetzt wurden. Die genauen Ansätze und PCR-Programm sind im Anhang zu finden (Primersequenzen: A.2.3 und Ansätze und PCR-Programm: A.7).

2.4.3.6 Auftrennung von RNA in Agarosegelen

RNA-Fragmente wurden, wie in 2.4.2.4 beschrieben, mittels Gelelektrophorese aufgetrennt.

2.4.3.7 Aufreinigung von PCR-Produkten

Zur Aufreinigung von RNA-PCR-Produkten wurde die unter 2.4.2.5 beschriebene Methode der Elution von PCR-Produkten aus Agarosegelen mit dem Promega-Kit durchgeführt.

2.4.3.8 Sequenzierung nach Sanger

Die Sequenzierung von RNA-PCR-Produkten erfolgte analog zu der Aufreinigung von DNA-PCR-Produkten (Abschnitt 2.4.2.6) mittels Kettenabbruchverfahren.

2.4.3.9 Quantitative Analyse des allelischen Transkriptverhältnisses

Zur Bestimmung des allelischen Transkriptverhältnisses wurde das SNaPshot-Verfahren benutzt. Hierbei handelt es sich im Prinzip um eine Sequenzierung nach Sanger, wobei jedoch ausschließlich ddNTPs (anstatt dNTPs) benutzt werden. Die Primer für die Reaktion liegen genau vor bzw. hinter dem zu untersuchenden SNP, so dass immer nur eine Base eingebaut werden kann. Die PCR-Produkte der RT-PCR wurden mittels Gelelektrophorese (2.4.2.4) überprüft, aufgereinigt (2.4.2.5) und am NanoDrop gemessen (2.4.2.2). Anschließend wurde die Konzentration der PCR-Produkte nach der folgenden Formel in pmol umgewandelt und Verdünnungen mit der Konzentration 0,1pmol/μl hergestellt:

$$\frac{1000}{650 * \text{Produktgröße in bp}} * \text{Konzentration ng}/\mu\text{l} = \text{pmol}/\mu\text{l}$$

Der SNaPshot-Ansatz mit einem Endvolumen von 10μl beinhaltet 0,2pmol PCR-Produkt, 1x SNaPshot®-Mix (Life Technologies) und 0,1μM Primer (for- oder rev-Primer). Die eigentliche SNaPshot-Reaktion fand mit folgenden Parametern im Thermocycler statt:

3min	96°C	(Denaturierung)	} 25 Zyklen
10sek	96°C	(Denaturierung)	
5sek	60°C	(Annealing)	
30sek	60°C	(Elongation)	
∞	4°C		

Anschließend wurde jedem Ansatz 1μl Shrimp Alkaline Phosphatase (SAP - 1U/μl; GE Healthcare) zur Aufreinigung hinzugegeben und bei 37°C für 45min im Thermocycler inkubiert. Die Reaktion wurde dann bei 75°C für 15min terminiert.

Anschließend wurde eine Fragmentanalyse (Genescan) durchgeführt, die zur Quantifizierung der Expression der unterschiedlichen Allele dient. Die aufgereinigten SNaPshot-Produkte wurden für den Genescan vorbereitet. Hierfür wurde jeweils ein 10μl Ansatz mit 1μl SNaPshot-Produkt, 0,5μl Genescan™-120LIZ™ Standard (Life Technologies) und 8,5μl Formamid hergestellt. Bevor der Genescan gestartet werden konnte, mussten die Produkte noch im Thermocycler denaturiert werden. Dies erfolgte bei 95°C für 5min. Der Genescan wurde auf dem Genetic Analyzer 3130 (Life Technologies) durchgeführt. Anschließend wurden die Daten mit dem Programm GeneMapper5 (Applied Biosystems) ausgewertet und normalisiert.

2.5 Methylierungsanalysen

2.5.1 Auswertung Illumina HiSeq2500

Auf Grund der großen Datenmenge und der limitierten Rechnerkapazität am Institut für Humangenetik, erfolgte die Datenverarbeitung der Methylohm-Rohdaten durch Herrn Christopher Schröder am Lehrstuhl für Genominformatik. Hierzu wurden im ersten Schritt noch vorhandene Adaptern der *paired-end* Reads mit dem Tool cutadapt (Parameter: Minimumlänge 30bp, Qualität 20) entfernt (Martin, 2011). Anschließend wurden die Reads mittels methylCtools (implementiert von Volker Hovestadt et al.; nicht veröffentlicht) mit Standardeinstellungen an die Referenz hs37d5 gemappt. MethylCtools benutzt zur Alignierung der Reads an das Referenzgenom das *Burrows-Wheeler Alignment Tool* (BWA) (Li u. Durbin, 2010). Mit SAMtools wurden die BAM-Dateien sortiert und mit Aufsummieren der Ausgabe von mpileup die *coverage* berechnet. (Li u. a., 2009). Read-Duplikate entfernte das Tool MarkDuplicates von Picard. Methylierungswerte wurden anschließend in BED-Dateien gespeichert und mit dem Skript bedGraphToBigWig in BIGWIG-Dateien transformiert. Einzelne Reads von potentiellen DMRs konnten nun mit Hilfe des *Integrative Genomic Viewer* (IGV) analysiert werden (Thorvaldsdottir u. a., 2013). Über eine Webseite, zur Verfügung gestellt durch Christopher Schröder, ist es möglich, die Methylierungswerte (Methylohm 1 und Methylohm 2) für einzelne CpG-Dinukleotide oder auch ganze Regionen innerhalb des Universitätsklinikums Essen abzurufen und in Textdateien zu speichern. Weitere Analysen konnten somit anhand dieser Daten mit der Programmiersprache Perl oder der Statistik-Software R durchgeführt werden (URLs siehe Anhang A.1; Perl-Skripte siehe CD).

2.5.2 Auswertung Roche/454 GS Junior

Bei der Sequenzierung von Bisulfit-behandelter DNA kann es häufiger zu Sequenzierungsfehlern kommen als bei unbehandelter DNA. Dies liegt an der Unterrepräsentation der Base Cytosin, da unmethylierte Cytosine durch die Bisulfitbehandlung der DNA in Thymin konvertiert wurden, während methylierte Cytosine, die nur in einem CpG-Kontext vorkommen, vor dieser Umwandlung geschützt sind. Bisulfit-Sequenzen neigen zur Bildung von (Thymin-) Homopolymeren. Daher sind die Qualitätswerte der Sequenzen deutlich niedriger, als bei unbehandelter DNA. Wie auch in der Roche Sequencing Application

Note beschrieben (Horsthemke u. a., 2013), wurden die Filtereigenschaften modifiziert (Herabsetzen der Qualitätsschwellenwerte), so dass die Anzahl an Reads deutlich erhöht werden konnte.

Die anschließende Auswertung der Methylierungsdaten erfolgte bei Standardeinstellungen mit der am Lehrstuhl für Genominformatik des Instituts für Humangenetik entwickelten Amplifyer-Software (Rahmann u. a., 2013).

3 Ergebnisse

Der Ergebnisteil dieser Arbeit unterteilt sich in zwei Abschnitte. Im ersten Teil (Abschnitt 3.1) werden die Ergebnisse der Analyse, ob es sich bei dem *RB1/Rb1*-Locus um einen einzigartigen Locus (siehe Abschnitte 1.6 und 1.7) in Bezug auf seine Eigenschaften handelt, dargestellt. Die Ergebnisse der genomweiten Analyse aller humanen CGIs, die mit einer Retrokopie überlappen (Abschnitt 3.2), werden im zweiten Abschnitt behandelt.

3.1 Ist der *RB1/Rb1*-Locus einzigartig in seinen Eigenschaften?

3.1.1 Identifikation von intronischen CGIs im Menschen, die es nicht in der Maus gibt

Um herauszufinden, ob der *RB1/Rb1*-Locus einzigartig in seinen Eigenschaften ist, wurden zuerst alle humanen (n=27.537) und murinen (n=15.997) CGIs auf ihre Lage in Relation zu proteinkodierenden Regionen (CCDS; n=14.990) untersucht (siehe Tabelle 3.1). Hierbei zeigte sich, dass nur eine kleine Anzahl CGIs ausschließlich in einem Intron liegt, so wie es bei CpG85 des *RB1*-Gens der Fall ist. Im Vergleich zum murinen Genom hat das humane Genom verhältnismäßig mehr intronische CGIs (7,7% zu 3,5%).

In Intron 2 des murinen *Rb1* Gens ist das CpG85 nicht vorhanden, daher wurden im nächsten Schritt die Sequenzen der 2.174 humanen intronischen CGIs mit denen der 579 murinen intronischen CGIs verglichen (siehe Abschnitt 2.2.3). Insgesamt wurden 2.033 humane intronische CGIs gefunden, die nicht in der Maus vorhanden sind. Im weiteren Verlauf dieser Arbeit werden diese CGIs als human/nicht-murin intronische CGIs bezeichnet. Statistische Analysen in Bezug auf Länge, GC-Gehalt, Anzahl von CpGs und der Rate von beobachteten zu erwarteten CpG-Dinukleotiden (*obs./exp. ratio*) wurden für diese 2.033 human/nicht-murin intronischen CGIs durchgeführt. Es wurden jedoch keine signifikanten Unterschiede zu anderen CGI-Klassen und auch keine signifikanten

Gemeinsamkeiten innerhalb dieser CGI-Klasse festgestellt (Daten nicht gezeigt). Auch die Maus besitzt intronische CGIs, die beim Menschen nicht vorhanden sind. Insgesamt wurden 470 intronische murine CGIs gefunden, die nicht im humanen Genom vorhanden sind. In dieser Arbeit werden die Eigenschaften dieser 470 murinen CGIs jedoch nicht weiter analysiert.

Tabelle 3.1: Lage aller humanen und murinen CGIs in Abhängigkeit von proteinkodierenden Regionen (CCDS)

CGIs	Human (hg19)		Murin (mm10)	
	(#)	(%)	(#)	(%)
CGIs (UCSC)	27.537	—	15.997	—
CGIs (analysiert)	28.396 ^a	100,00	16.643 ^a	100,00
CGIs <i>intragenic</i>	17.807	62,71	12.587	75,63
CGIs <i>intergenic</i>	10.589	37,29	4.056	24,37
TSS	503	1,77	474	2,85
TSS + 5'UTR	3.412	12,02	3.029	18,20
TSS + 5'UTR + CDS(Exon)	1.078	3,80	1.189	7,14
TSS + 5'UTR + CDS(Exon) + CDS(Intron)	5.223	18,39	4.692	28,19
TSS + 5'UTR + CDS(Exon) + CDS(Intron) + 3'UTR	275	0,97	107	0,64
5'UTR	491	1,73	303	1,82
5'UTR + CDS(Exon)	93	0,33	112	0,67
5'UTR + CDS(Exon) + CDS(Intron)	273	0,96	225	1,35
5'UTR + CDS(Exon) + CDS(Intron) + 3'UTR	27	0,10	13	0,08
CDS(Exon)	845	2,98	781	4,69
CDS(Exon) + CDS(Intron)	2.453	8,64	758	4,55
CDS(Exon) + CDS(Intron) + 3'UTR	620	2,18	222	1,33
CDS(Exon) + 3'UTR	70	0,25	39	0,23
CDS(Intron)	2.174	7,66	579	3,48
3'UTR	270	0,95	64	0,38

Die Tabelle zeigt die Verteilung von humanen und murinen CGIs in Abhängigkeit von ihrer Lage zu proteinkodierenden Regionen. Zusätzlich zu der Anzahl (#) ist auch der prozentuale Anteil (%) angegeben. Es wurden fünf Klassen für die Einteilung von CGIs definiert (siehe auch Abschnitt 2.1). Ein CGI kann mit keiner, einer oder mehreren Klassen überlappen. Diese Tabelle wurde bereits in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

^a Die Anzahl analysierter CGIs ist höher, als die des UCSC Browser, da ein CGI mehreren CCDS zugeordnet werden kann (siehe auch Abschnitt 2.2.1).

Die Retrotransposition des *PPP1R26*-Gens, u.a. in das *RB1*-Gen, fand mehrfach im Laufe der Evolution statt. Um ein oder mehrere ähnliche Ereignisse zu finden, wurde mit den Sequenzen der 2.033 human/nicht-murin intronischen CGIs eine MEGA-Blast-Suche durchgeführt (siehe Abschnitt 2.2.3). Bei 135 CGIs wurden Sequenzähnlichkeiten zu einer

oder mehreren Regionen im menschlichen Genom gefunden. Jedoch wurden 31 dieser 135 CGIs von weiteren Analysen ausgeschlossen, da sie entweder eine sehr geringe Sequenzähnlichkeit aufwiesen (<25bp), identisch zu ihrer eigenen Sequenz oder aber identisch zu ihrem ancestralen Gen sind. Somit blieben 104 human/nicht-murin intronische CGIs für weitere Analysen übrig. Die meisten zusätzlichen Treffer (n=76) der 104 human/nicht-murin CGIs überlappen mit Transkriptionsstartstellen (TSS), kodierenden Bereichen (CDS) von Genen oder beidem (TSS und CDS). Nur 13 CGIs weisen einen zusätzlichen Treffer in einem Intron auf und 15 CGIs einen Treffer außerhalb von proteinkodierenden Regionen. Im nächsten Schritt wurde überprüft, ob die 104 human/nicht-murin intronischen CGIs mit einer Retrokopie überlappen, so wie CpG85 mit dem *retro-PPP1R26* überlappt. Bei 45 CGIs war dies der Fall. Sogar 20 CGIs wiesen mindestens zwei zusätzliche Treffer mit einer Retrokopie auf. Die übrigen 59 CGIs konnten nicht mit Retrokopien in Verbindung gebracht werden. Daher liegt die Vermutung nahe, dass sie eher mit einer nicht-annotierten Retrokopie oder nur einem sehr kleinen Abschnitt einer Retrokopie (der nicht als Retrokopie identifiziert wird) überlappen, als dass die CGIs durch Genduplikation entstanden sind.

3.1.2 Methylierungsanalysen von human/nicht-murin intronischen CGIs

Der Methylierungsgrad der 104 human/nicht-murin intronischen CGIs, die eine hohe Sequenzähnlichkeit zu anderen Regionen im menschlichen Genom aufweisen, wurde anhand von zwei Monozyten-Methylomen (Methylom 1: 1.929.952.791 Reads, Duplikatraterate 0,22, Mappingrate 0,99, Bisulfitkonvertierungsrate 0,994; Methylom 2: 1.407.767.072 Reads, Duplikatraterate 0,15, Mappingrate 0,98, Bisulfitkonvertierungsrate 0,996) analysiert (siehe Tabelle 3.2). Der Tabelle 3.2 ist zu entnehmen, dass ca. 75% der human/nicht-murin intronischen CGIs methyliert sind. Nur 15 CGIs weisen eine Methylierung unter 20% auf (unmethyliert) und nur 12 CGIs sind zwischen 20% und 80% methyliert (Methylom 1). Beide Methylome zeigen ähnliche Ergebnisse, wobei die genomische Abdeckung (*coverage*) von Methylom 2 deutlich schlechter ist. Für weitere Analysen wurden 14 CGIs ausgewählt, die in mindestens einem Methylom eine mittlere Methylierung zwischen 20% und 80% aufwiesen. Neun von diesen 14 CGIs zeigen in beiden Methylomen eine mittlere Methylierung. Drei CGIs weisen ausschließlich in Methylom 1 eine Methylierung zwischen 20% und 80% auf und in Methylom 2 sind es zwei CGIs.

Tabelle 3.2: Methylierungsanalyse von 104 human/nicht-murin intronischen CGIs

		Methylierung ($m < 20\%$)	Methylierung ($20\% \geq m \leq 80\%$)	Methylierung ($m > 80\%$)
Methylom 1				
Anzahl CGIs (#)	104	15	12	77
Ø Methylierung (%)	77	2	54	95
Ø Abdeckung	14	8	12	12
Methylom 2				
Anzahl CGIs (#)	104	18	11	75
Ø Methylierung (%)	77	4	68	95
Ø Abdeckung	8	5	8	9

Die Tabelle zeigt eine Übersicht über den Grad der Methylierung und die Anzahl analysierter CGIs in zwei Monozyten-Methylomen. Die CGIs sind nach ihrer Methylierung (m) in drei Klassen unterteilt: unmethyliert ($m < 20\%$), methyliert ($m > 80\%$) und CGIs mit einer Methylierung zwischen methyliert und unmethyliert ($20\% \geq m \leq 80\%$). Zusätzlich ist noch die Abdeckung (*coverage*) der 104 analysierten CGIs für jedes Methylom angegeben. Diese Tabelle wurde in abgeänderter Form bereits in Rademacher u. a. (2014) veröffentlicht.

3.1.2.1 Suche nach allelspezifischer Methylierung

Im nächsten Schritt wurde überprüft, ob die mittlere Methylierung der 14 CGIs aufgrund einer allelspezifischen Methylierung vorliegt. Hierfür wurde zunächst der Methylierungsgrad einzelner Reads genau analysiert (siehe Tabelle 3.3, Abbildung 3.1 und Abbildungen A.1, A.2, A.3 und A.4 im Anhang A.4.2). Bei einem differentiell methylierten CGI werden hauptsächlich methylierte ($m > 80\%$) bzw. unmethylierte Reads ($m < 20\%$) erwartet und nur ein geringer Anteil an mix-methylierten Reads ($20\% \geq m \leq 80\%$). Um CGIs auszuschließen, die einen vollständig methylierten Bereich und einen vollständig unmethylierten Bereich haben (siehe Abb. 3.1 (B)), wurde der Methylierungsgrad aller CpG-Dinukleotide eines CGIs untersucht. Hierfür wurden Varianz (VAR) und Standardabweichung (SD) berechnet (siehe Spalte VAR und SD in Tab. 3.3). Die Standardabweichung gibt an, wie groß die Abweichung vom Mittelwert ist. Die Methylierung einzelner CpGs sollte bei einem differentiell methylierten CGI bei ca. 50% liegen und die Standardabweichung möglichst gering sein (siehe Abbildung 3.1 (A)).

Zufällig methylierte CGIs können vollständig methylierte und vollständig unmethylierte CpGs haben und dadurch eine hohe Standardabweichung (siehe Abbildung 3.1 (B) und (C)). Basierend auf diesen Kriterien (Anteil partiell methylierter Reads $\leq 20\%$ und VAR $\leq 0,05$) wurden neun CGIs von den weiteren Analysen ausgeschlossen.

Tabelle 3.3: Methylierungsanalysen einzelner Reads und CpG-Dinukleotide der CGIs mit einem mittleren Methylierungslevel (Methylom 1 und Methylom 2)

CGI ID	Gen	Chr.	Länge (bp)	\emptyset Methylierung (%)	\emptyset Abdeckung	# Reads	unmeth. Reads (#)	(%)	meth. Reads (#)	(%)	part. meth. Reads (#)	(%)	CpG Methylierung VAR	SD
Methylom 1														
1911_1_hg19	<i>DCAF</i>	1	214	81	18	70	2	3	48	69	20	29	0.01	0.11
4675_1_hg19	<i>GXYLT2</i>	3	326	73	19	108	8	7	70	65	30	28	0.02	0.12
4754_1_hg19	<i>SLC9CI</i>	3	463	32	9	66	31	47	15	23	20	30	0.09	0.29
9009_1_hg19	<i>MAD1L1</i>	7	210	78	8	24	2	8	14	58	8	33	0.08	0.29
14414_1_hg19	<i>ASRGL1</i>	11	252	55	10	40	17	43	21	53	2	5	0.02	0.13
15205_1_hg19	<i>CACNA2D4</i>	12	1219	84	12	188	7	4	143	76	38	20	0.03	0.18
15224_1_hg19	<i>PARP11</i>	12	698	42	11	108	53	49	45	42	10	9	0.05	0.21
15290_1_hg19	<i>CDI63L1</i>	12	624	23	14	153	57	37	41	27	55	36	0.08	0.27
16634_1_hg19	<i>RBI</i>	13	1222	63	13	193	69	36	116	60	8	4	0.02	0.13
19100_1_hg19	<i>PDXDC1</i>	16	679	75	11	104	22	21	67	64	15	14	0.02	0.16
19870_1_hg19	<i>SLC7A5</i>	16	207	58	15	45	17	38	17	38	11	24	0.19	0.44
20632_1_hg19	<i>MYO1D</i>	17	466	22	7	54	38	70	11	20	5	9	0.03	0.18
20636_1_hg19	<i>ASIC2</i>	17	506	58	6	44	4	9	20	45	20	45	0.08	0.29
25767_1_hg19	<i>HSP2BP</i>	21	403	70	17	93	22	24	50	54	21	23	0.10	0.32
Methylom 2														
1911_1_hg19	<i>DCAF</i>	1	214	58	6	25	4	16	9	36	12	48	0.08	0.29
4675_1_hg19	<i>GXYLT2</i>	3	326	79	15	69	2	3	48	70	19	28	0.02	0.15
4754_1_hg19	<i>SLC9CI</i>	3	463	17	6	49	24	49	7	14	18	37	0.05	0.23
9009_1_hg19	<i>MAD1L1</i>	7	210	77	2	7	0	0	4	57	3	43	0.14	0.37
14414_1_hg19	<i>ASRGL1</i>	11	252	49	9	36	16	44	17	47	3	8	0.02	0.14
15205_1_hg19	<i>CACNA2D4</i>	12	1219	76	6	95	7	7	68	72	20	21	0.08	0.28
15224_1_hg19	<i>PARP11</i>	12	698	46	8	87	35	40	45	52	7	8	0.11	0.33
15290_1_hg19	<i>CDI63L1</i>	12	624	20	11	114	48	42	30	26	36	32	0.05	0.22
16634_1_hg19	<i>RBI</i>	13	1222	72	8	107	36	34	66	62	5	5	0.03	0.17
19100_1_hg19	<i>PDXDC1</i>	16	679	79	6	48	9	19	37	77	2	4	0.04	0.21
19870_1_hg19	<i>SLC7A5</i>	16	207	59	8	26	9	35	8	31	9	35	0.20	0.45
20632_1_hg19	<i>MYO1D</i>	17	466	2	2	13	13	100	0	0	0	0	0.02	0.14
20636_1_hg19	<i>ASIC2</i>	17	506	74	5	37	2	5	22	59	13	35	0.07	0.26
25767_1_hg19	<i>HSP2BP</i>	21	403	75	13	73	18	25	34	47	21	29	0.10	0.32

Diese Tabelle zeigt die Ergebnisse der Methylierungsanalyse von einzelnen Reads und von CpG-Dinukleotiden der 14 potentiell differentiell methylierten CGIs für Methylom 1 und Methylom 2 (*coverage* von Methylom 2 deutlich schlechter als die von Methylom 1). Angegeben sind die Namen der CGIs (CGI ID), sowie Gen, Chromosom, Länge, Durchschnittsmethylierung, Durchschnittsabweichung (*coverage*) und die Anzahl analysierter Reads. Zusätzlich sind die einzelnen Reads in drei Klassen unterteilt: Unmethylierte ($m < 20\%$), methylierte ($m > 80\%$) und partiell methylierte Reads ($20\% \geq m \leq 80\%$). Die letzte Spalte zeigt die Ergebnisse der Methylierungsanalyse der CpG-Dinukleotide. Varianz (VAR) und Standardabweichung aller einzelner CpGs eines CGIs wurden berechnet. Fett: Kandidaten-CGIs, die differentiell methyliert sein könnten (Kriterien: Anteil partiell methylierter Reads $\leq 20\%$ und VAR $\leq 0,05$). Die Tabelle ist in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht worden.

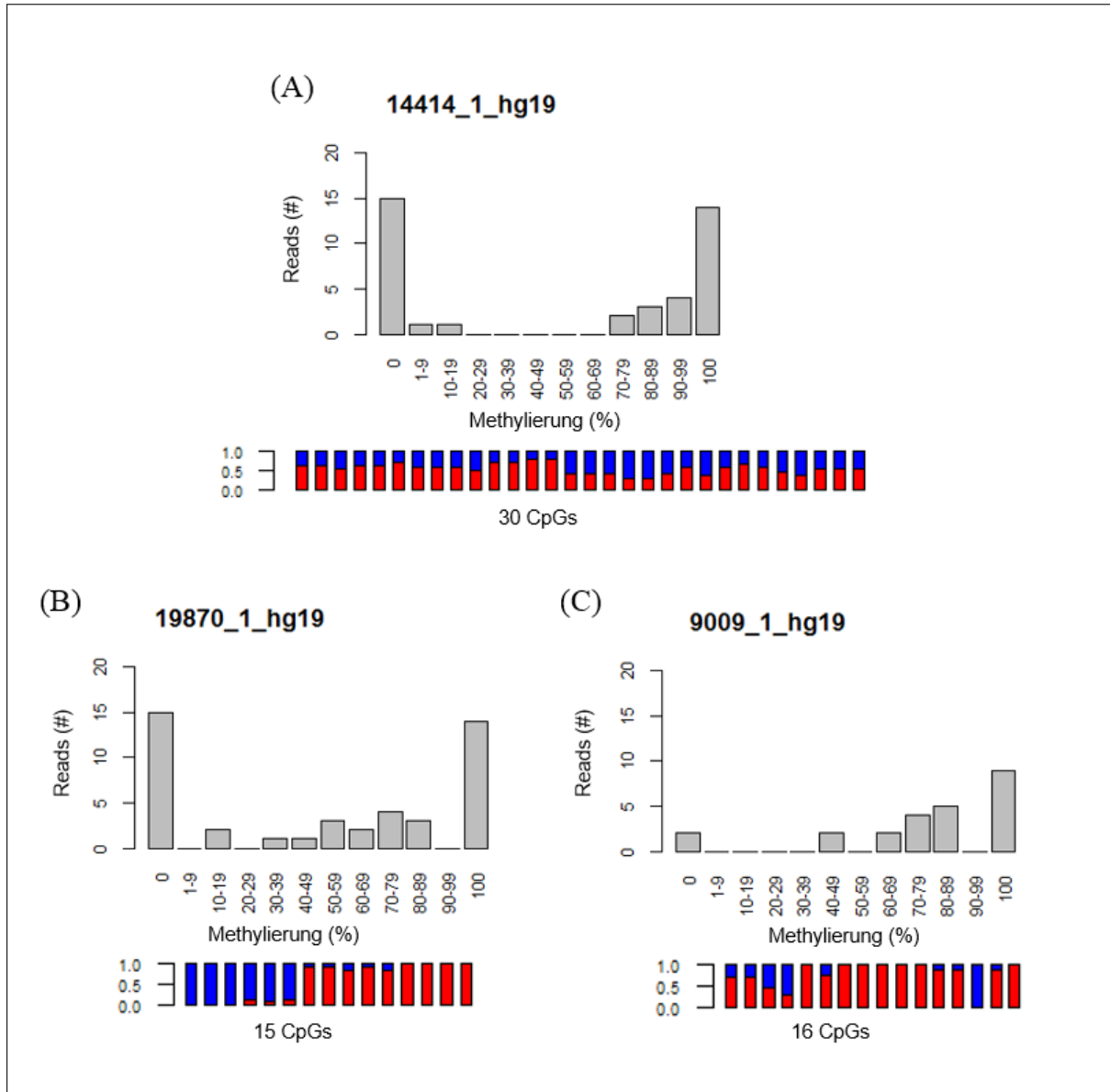


Abbildung 3.1: Methylierungsmuster von drei human/nicht-murin intronischen CGIs mit einem Methylierungsgrad zwischen 20% und 80%

Für jedes CGI zeigt das Histogramm die Verteilung von Reads (y-Achse) in Abhängigkeit von ihrer Methylierung (x-Achse). Die Balkendiagramme unter dem Histogramm zeigen die Methylierung der einzelnen CpG-Dinukleotide innerhalb des CGIs, wobei methylierte CpGs rot und unmethylierte CpGs blau dargestellt sind. **(A)** zeigt das CGI 14414_1_hg19 (*ASRGL1*) mit hauptsächlich methylierten und unmethylierten Reads. Jedes einzelne CpG besitzt eine Methylierung von ca. 50%. Dieses Ergebnis zeigt, dass das CGI differentiell methyliert sein könnte. **(B)** Auch das CGI 19870_1_hg19 (*SLC7A5*) zeigt im Histogramm hauptsächlich methylierte und unmethylierte Reads. Im Balkendiagramm wird jedoch deutlich, dass dieses CGI aus einem unmethylierten 5'-Ende und einem methylierten 3'-Ende besteht. Dieses Ergebnis zeigt, dass keine differentielle Methylierung vorliegt. **(C)** CGI 9009_1_hg19 (*MAD1L1*) zeigt weder eine bimodale Verteilung im Histogramm noch eine Methylierung von 50% bei jedem einzelnen CpG-Dinukleotid. Dieses CGI ist somit nicht differentiell methyliert (bereits veröffentlicht in abgeänderter Form in Rademacher u. a. (2014)).

Die restlichen fünf CGIs sind mit den folgenden Genen assoziiert: *ASRGL1* (14414_1_hg19), *PARP11* (15224_1_hg19), *RB1* (16634_1_hg19), *PDXDC1* (19100_1_hg19) und *MYO1D* (20632_1_hg19) (Screenshots einzelner Reads dieser fünf CGIs sind auf der Daten-CD zu finden).

Im nächsten Schritt wurde das Methylierungsniveau dieser fünf CGIs in bereits publizierten Methylogen (aus Ziller u. a. (2013); unterschiedliche Gewebetypen) untersucht (siehe Tabelle 3.4). Abgesehen von dem Methylierungsgrad in DNA aus humanen Spermien, in der alle fünf CGIs unmethyliert sind, wurde in fast allen Gewebetypen eine Methylierung zwischen 20% und 80% beobachtet. Gewebespezifische Unterschiede konnten in den unterschiedlichen Methylogen detektiert werden. Zum Beispiel ist in fetalen Geweben die Methylierung geringer als in adulten Geweben. Leider sind von diesen Datensätzen keine Readdaten verfügbar. Somit war eine Methylierungsanalyse wie in Tabelle 3.3 nicht möglich.

3.1.2.2 Methylierungsanalysen mittels Roche/454 GS Junior

Mit Hilfe der hochquantitativen Hochdurchsatz-Bisulfitsequenzierung auf dem Roche/454 GS Junior wurde untersucht, ob das mittlere Methylierungsniveau der fünf human/nicht-murin intronischen CGIs auf eine allelspezifische Methylierung zurückzuführen ist, wie es bei dem *RB1*-Locus der Fall ist (Kanber u. a., 2009b). Es wurden Amplikons für die Bisulfitsequenzierung von Bisulfit-behandelter DNA aus Monozyten von unverwandten Spendern, die heterozygot für einen bestimmten SNP sind, generiert. Diese wurden anschließend auf dem Roche/454 GS Junior sequenziert. Für drei Loci (*ASRGL1*, *PARP11*, *PDXDC1*) konnten Amplikons generiert werden, jedoch war dies nicht für den *MYO1D*-Locus möglich. Insgesamt wurde die DNA von 22 Spendern auf Heterozygotie für den jeweiligen SNP der Region überprüft. Für jeden Locus wurden mindestens 12 informative Proben ermittelt. Für den *ASRGL1*-Locus waren 14 Proben heterozygot für einen A/G SNP (rs11231058). Vier Individuen zeigten deutliche (> 10%) Methylierungsunterschiede der Allele (Beispiel siehe Abbildung 3.2 (A); Tabelle A.11 siehe Anhang A.4.3). Bei drei Individuen war das Allel G weniger methyliert und bei einem Individuum das Allel A. Die übrigen 10 Proben zeigten keine Unterschiede in der Methylierung der Allele. Diese Ergebnisse können darauf hinweisen, dass es sich um eine zufällige Variation oder aber einen elternspezifischen Effekt handelt. Insgesamt zwölf Spender waren heterozygot für einen A/G SNP (rs12319851) im *PARP11*-Locus. In den meisten Fällen (11 von 12) waren die Methylierungsunterschiede zwischen den beiden Allelen geringer als 10% (Beispiel siehe

Tabelle 3.4: Methylierungsgrad der fünf potentiell differentiell methylierten CGIs in anderen Geweben

Probenname	Zellen/Gewebe		14414_1_hg19 (ASRGL1)		15224_1_hg19 (PARP11)		16634_1_hg19 (RBI)		19100_1_hg19 (PDXDC1)		20632_1_hg19 (MYO1D)	
			Meth.(%)	Abd.	Meth.(%)	Abd.	Meth.(%)	Abd.	Meth.(%)	Abd.	Meth.(%)	Abd.
Monozyten Methylohm 1 ^a	Monozyten		55	10	42	11	63	13	75	11	22	7
Monozyten Methylohm 2 ^a	Monozyten		49	9	46	8	72	8	79	6	2	2
Frontal_cortex_normal_1 ^b	Kortex		84	27	53	38	76	48	28	57	4	33
Frontal_cortex_normal_2 ^b	Kortex		83	20	63	33	76	32	42	34	13	24
Frontal_cortex_AD_1 ^b	Kortex		89	22	52	30	76	38	39	45	14	33
Frontal_cortex_AD_2 ^b	Kortex		88	36	69	43	76	56	41	40	16	44
IMR90 ^b	Lunge, fetal, Fibroblasten		56	18	53	13	82	19	53	32	25	11
Colon_Primary_Normal ^b	Darm		52	38	67	42	79	59	44	47	12	23
Human sperm	Sperma ^b		9	4	1	5	5	7	7	11	6	16
Adult liver replicate 1 ^b	Leber		82	61	81	74	71	77	54	50	10	28
Adult liver replicate 2 ^b	Leber		56	53	52	45	61	57	47	49	11	38
Hippocampus middle replicate 1 ^b	Hippokampus (Zentrum)		88	62	67	50	77	65	36	55	19	26
Hippocampus middle replicate 2 ^b	Hippokampus (Zentrum)		87	35	69	39	77	48	30	46	16	26
Fetal heart (119) ^b	fetales Herz		64	39	38	37	58	51	34	53	7	27
Fetal thymus (1238) ^b	fetaler Thymus		65	29	29	43	49	42	34	63	22	36
Fetal adrenal (1244) ^b	fetale Nebenniere		52	33	37	26	67	29	52	44	10	26
Fetal muscle leg (1243) ^b	fetaler Muskel (Bein)		55	35	28	44	54	38	32	52	5	29
Fetal brain (515) ^b	fetales Gehirn		78	22	46	21	51	30	19	34	10	20

Die Tabelle zeigt die Methylierungslevel der CGIs 14414_1_hg19 (*ASRGL1*), 15224_1_hg19 (*PARP11*), 16634_1_hg19 (*RBI*), 19100_1_hg19 (*PDXDC1*) und 20632_1_hg19 (*MYO1D*) in unterschiedlichen Gewebetypen. Angegeben ist jeweils neben dem durchschnittlichen Methylierungsgrad (Meth.) in Prozent die durchschnittliche Abdeckung (Abd.) für das jeweilige CGI. Diese Tabelle wurde bereits in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

^a veröffentlicht in Rademacher u. a. (2014)

^b veröffentlicht in Ziller u. a. (2013)

Abbildung 3.2 (C); Tabelle A.13 siehe Anhang A.4.3). Für den *PDXDC1*-Locus waren insgesamt 14 der 22 Spender für einen A/C SNP heterozygot (rs9928601). Fast alle Individuen zeigten Methylierungsunterschiede zwischen den beiden Allelen (im Durchschnitt um 40%). In 13 Fällen war das C-Allel das weniger methylierte.

Parallel zu den Analysen mit Monozyten-DNA wurde DNA aus Blut von 20 Individuen (siehe Abschnitt 2.3.6.2) untersucht, von denen auch elterliches Material zur Verfügung stand. Im ersten Schritt wurde untersucht, ob diese heterozygot für die ausgewählten SNPs sind. Wenn dies der Fall war, wurde die DNA der Eltern analysiert, um informative Trios zu identifizieren (mindestens ein Elternteil homozygot für ein Allel). Es wurden insgesamt sechs informative Trios identifiziert (zwei für den *ASRGL1*-Locus, drei für den *PARP11*-Locus und eins für den *PDXDC1*-Locus) und Amplikons zur Analyse auf dem Roche/454 GS Junior generiert. Für den *PDXDC1*-Locus wurden zusätzlich noch drei heterozygote Individuen untersucht, bei denen jedoch eine Bestimmung der elterlichen Herkunft der Allele nicht möglich ist, da beide Elternteile ebenfalls heterozygot für den SNP sind.

Bei der Methylierung des *ASRGL1*-Locus konnten in einem Beispiel keine Unterschiede in der Methylierung der Allele identifiziert werden und im anderen eine geringere Methylierung des G-Allels (siehe Abbildung 3.3 (A) und ausführliche Tabelle A.14 im Anhang A.4.4). Alle drei untersuchten Proben für den *PARP11*-Locus zeigten keine Unterschiede in der Methylierung der einzelnen Allele (Abbildung 3.3 (B); Tabelle A.14). Einen deutlichen Unterschied in der Methylierung zeigten zwei Proben des *PDXDC1*-Locus (Abbildung 3.3 (C); Tabelle A.14). In beiden Fällen war das C-Allel geringer methyliert, wobei man in dem dargestellten Fall in Abbildung 3.3 von einem unmethylierten C-Allel ($m < 20\%$) und einem methylierten A-Allel ($m > 80\%$) sprechen kann. Das unmethylierte C-Allel wurde in diesem Fall paternal vererbt. Eine Aussage über den elterlichen Ursprung der Allele konnte im zweiten Fall nicht gemacht werden, da alle Mitglieder des Trios heterozygot für den untersuchten SNP sind (siehe Anhang A.4.4 Abbildung A.12 für weitere Methylierungsplots inklusive der Stammbäume).

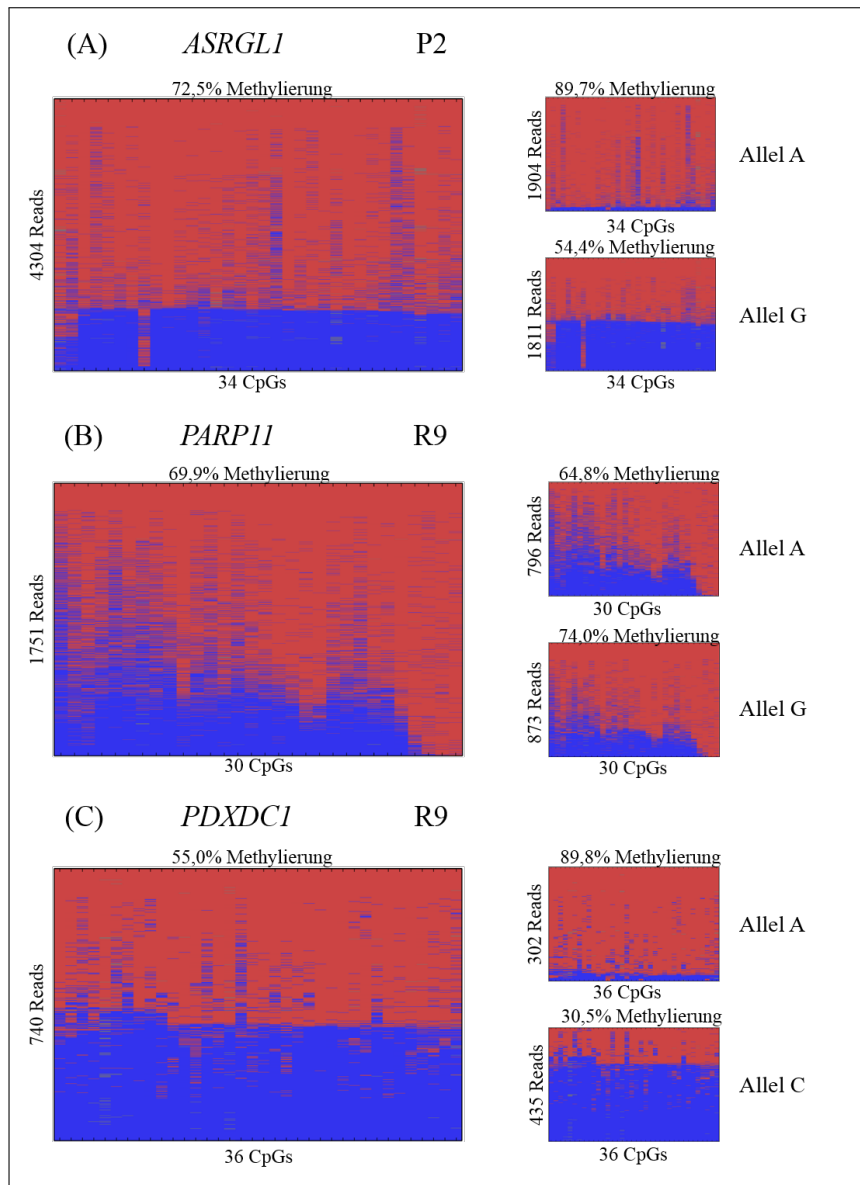


Abbildung 3.2: Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozygoten Individuen

Die Abbildung zeigt für jeden untersuchten Locus beispielhaft ein Ergebnis einer Methylierungsanalyse, wobei die Amplikons jeweils nur einen Teil der CGIs abdecken bzw. noch flankierende CpGs einschließen können. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons, die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. Über jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben. Links steht die Anzahl an Reads und unter dem Plot die Anzahl analysierter CpGs. P2 und R9 sind die institutsinternen Namen der Monozytenproben. Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. (A) zeigt das Methylierungsmuster von CGI 14414_1_hg19 (*ASRGL1*), wobei die ersten zwei CpGs nicht zum CGI gehören, (B) das Methylierungsmuster von CGI 15224_1_hg19 (*PARP11*) und (C) das Methylierungsmuster von CGI 19100_1_hg19 (*PDXDC1*), wobei die letzten vier CpGs nicht zum CGI gehören. Diese Abbildung wurde bereits in abgeänderter Form veröffentlicht in Rademacher u. a. (2014).

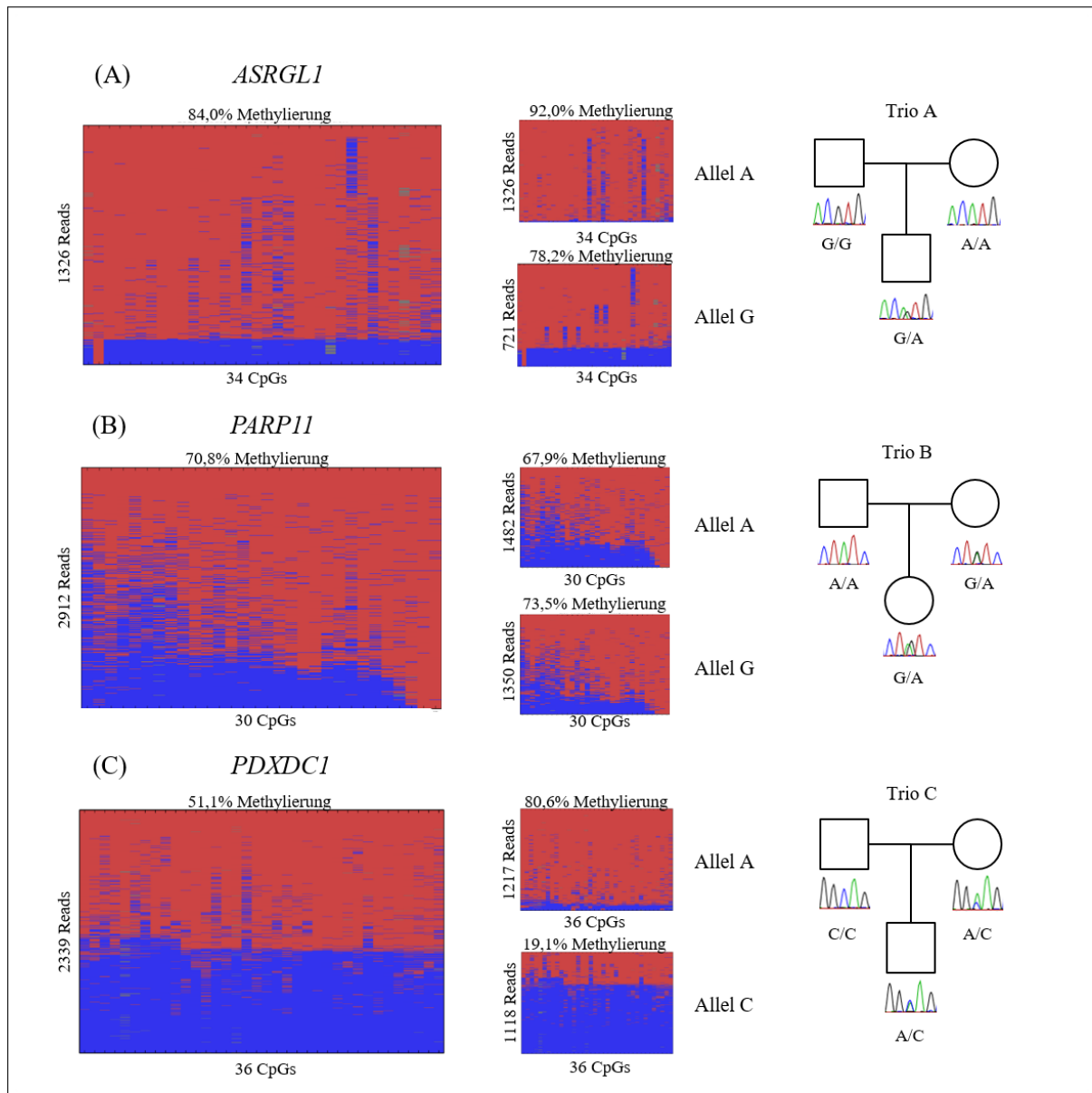


Abbildung 3.3: Methylierungsanalyse einzelner CpGs der drei intronischen CGIIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Blut von heterozygoten Individuen, die informativ in Bezug auf die elterliche Herkunft sind

Die Abbildung zeigt für jeden untersuchten Locus beispielhaft ein Ergebnis der Methylierungsanalyse, wobei die Amplikons jeweils nur einen Teil der CGIIs abdecken bzw. noch flankierende CpGs einschließen können. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons, die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. Über jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben. Links steht die Anzahl an Reads und unter dem Plot die Anzahl analysierter CpGs. Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Rechts neben den Plots ist der Stammbaum der jeweiligen Familie angegeben, der die Elektropherogramme des analysierten SNPs für die Familienmitglieder enthält. (A) zeigt das Methylierungsmuster von CGI 14414_1_hg19 (*ASRGL1*), wobei die ersten zwei CpGs nicht zum CGI gehören, (B) das Methylierungsmuster von CGI 15224_1_hg19 (*PARP11*). und (C) das Methylierungsmuster von CGI 19100_1_hg19 (*PDXDC1*), wobei die letzten vier CpGs nicht zum CGI gehören. Das unmethylierte C-Allel wurde in diesem Fall paternal vererbt.

3.1.3 Evolutionärer Ursprung von human/nicht-murin intronischen CGIs

Der evolutionäre Ursprung von human/nicht-murin intronischen CGIs wurde mit Hilfe von BLAST-Suchen in sieben Primaten analysiert (siehe Abschnitt 2.2.4). Hierbei wurde untersucht, ob die Sequenz des humanen intronischen CGIs und Sequenzen von flankierenden humanen Exons in dem jeweiligen Primaten-Referenzgenom vorhanden sind. Die Analysen wurden bei den 104 human/nicht-murin intronischen CGIs durchgeführt, die zusätzlich eine hohe Sequenzähnlichkeit zu anderen Regionen im menschlichen Genom aufwiesen und bei denen die Methylierungsanalysen durchgeführt wurden. Bei 86 von 104 CGIs war es möglich, den Zeitpunkt der Entstehung während der Evolution näher zu bestimmen (siehe Tabelle 3.5 und Tabelle A.23 im Anhang). Aufgrund von Sequenzierungslücken in einigen untersuchten Genomen konnten nicht alle Zeitpunkte bestimmt werden.

Tabelle 3.5: Evolutionäre Analyse von 86 human/nicht-murin intronischen CGIs in unterschiedlichen Primaten

Unterordnung Teilordnung Überfamilie	Haplorrhini							Strepsirrhini
	Catarrhini					Platyrrhini		
	Hominoidea			Cercopithecoidea				
Organismus	Mensch	Schimpanse	Gorilla	Orang-Utan	Gibbon	Rhesusaffe	Marmosette	Galago
CGIs (#)	86	73	67	59	59	44	29	7

Die Tabelle gibt einen Überblick über die Evolution von 86 human/nicht-murin intronischen CGIs. Die Anzahl an CGIs gibt an, wie viele CGIs in dem untersuchten Organismus vorhanden sind. Zusätzlich sind noch die Unterordnungen, Teilordnungen und Überfamilien der untersuchten Primaten angegeben. Diese Tabelle wurde bereits in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

Mehr als die Hälfte der human/nicht-murin intronischen CGIs (57%) sind in allen untersuchten Mitgliedern der Überfamilie Hominoidea (Mensch, Schimpanse, Gorilla, Orang-Utan und Gibbon) vorhanden. Nur sieben der human/nicht-murin intronischen CGIs konnten im Galago-Genom (Buschbaby) nachgewiesen werden. Von den sieben untersuchten Primaten gehören nur die Galagos der Untergruppe Strepsirrhini an. Alle anderen untersuchten Primaten gehören zur Unterordnung der Haplorrhini (Perelman u. a., 2011).

Für alle fünf human/nicht-murin intronischen CGIs mit einem mittleren Methylierungslevel konnte der Zeitpunkt der Entstehung näher bestimmt werden (Abbildung 3.4). Wie auch schon in vorherigen Studien beschrieben wurde, ist das CGI 16634_1_hg19 (CpG85) des *RB1*-Gens in allen analysierten Mitgliedern der Haplorrhini vorhanden. Nicht vorhanden ist dieses CGI in Galago, die einer anderen Unterordnung (Strepsirrhini) angehören (Kanber

u. a., 2013). Die human/nicht-murin intronischen CGIs der Gene *ASRGL1* (14414_1_hg19) und *PARP11* (15224_1_hg19) sind in allen analysierten Mitgliedern der Überfamilie Hominoidea vorhanden. Das CGI 19100_1_hg19 (*PDXDC1*) existiert nur im Menschen, wohingegen das CGI 20632_1_hg19 (*MYO1D*) zusätzlich noch im Schimpanse vorhanden ist.

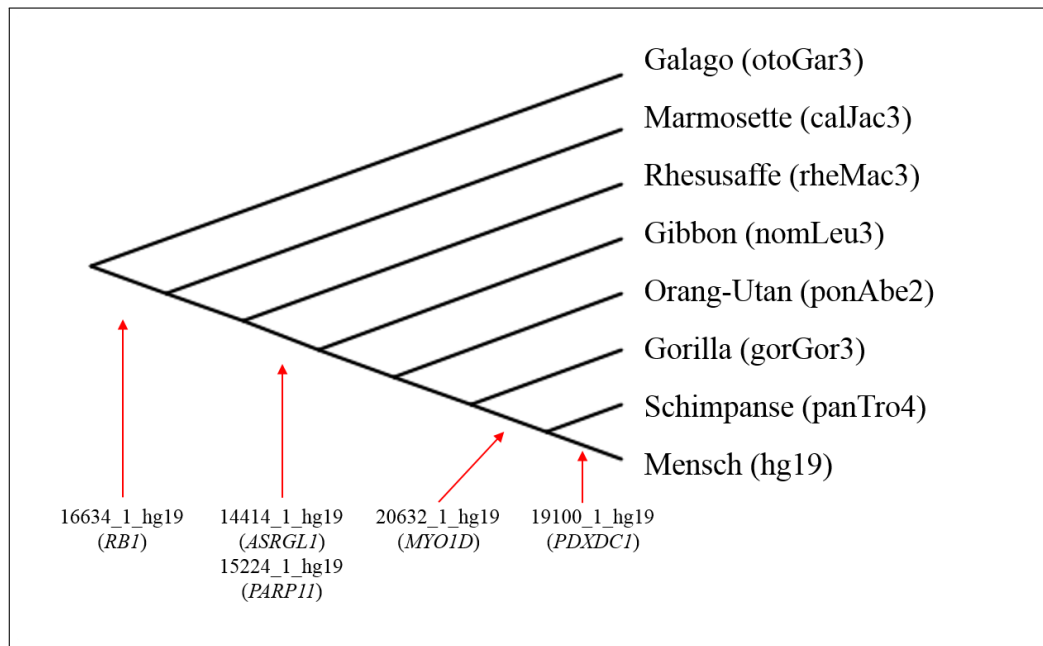


Abbildung 3.4: Entstehungszeitpunkte der fünf human/nicht-murin intronischen CGIs mit einem mittleren Methylierungsgrad

Die Abbildung zeigt einen vereinfachten Stammbaum (ohne Skalierung der Distanzen) aller sieben analysierten Primaten (Mensch, Schimpanse, Gorilla, Orang-Utan, Gibbon, Rhesusaffe, Marmosette und Galago) inklusive der Version des Referenzgenoms. Die roten Linien zeigen die Zeitpunkte, an denen die CGIs entstanden sind (bereits veröffentlicht in abgeänderter Form in Rademacher u. a. (2014)).

Das CGI 16634_1_hg19 (*RBI*) ist Teil einer Retrokopie. Zusätzlich zu diesem CGI wurden drei human/nicht-murin intronische CGIs (14414_1_hg19 (*ASRGL1*), 15224_1_hg19 (*PARP11*) und 19100_1_hg19 (*PDXDC1*)) identifiziert, die ebenfalls Teil von Retrokopien sind (Abbildung 3.5). CGI 14414_1_hg19 (*ASRGL1*) ist Teil einer Retrokopie des *RCC2*-Gens auf Chromosom 11, CGI 15224_1_hg19 (*PARP11*) des *OTUD4*-Gens auf Chromosom 4 und CGI 19100_1_hg19 (*PDXDC1*) des *KIAA2013*-Gens auf Chromosom 1. Das CGI 16634_1_hg19 weist Sequenzähnlichkeiten zu zwei kleinen unmethylierten CGIs innerhalb des offenen Leserasters in Exon 4 des ancestralen Gens (*PPP1R26*) auf. Im Gegensatz hierzu haben die CGIs 14414_1_hg19 (*ASRGL1*), 15224_1_hg19 (*PARP11*)

und 19100_1_hg19 (*PDXDC1*) Sequenzähnlichkeiten mit unmethylierten CGIs, die das 5'-Ende der ancestralen Gene (*RCC2*, *OTUD4* und *KIAA2013*) überspannen.

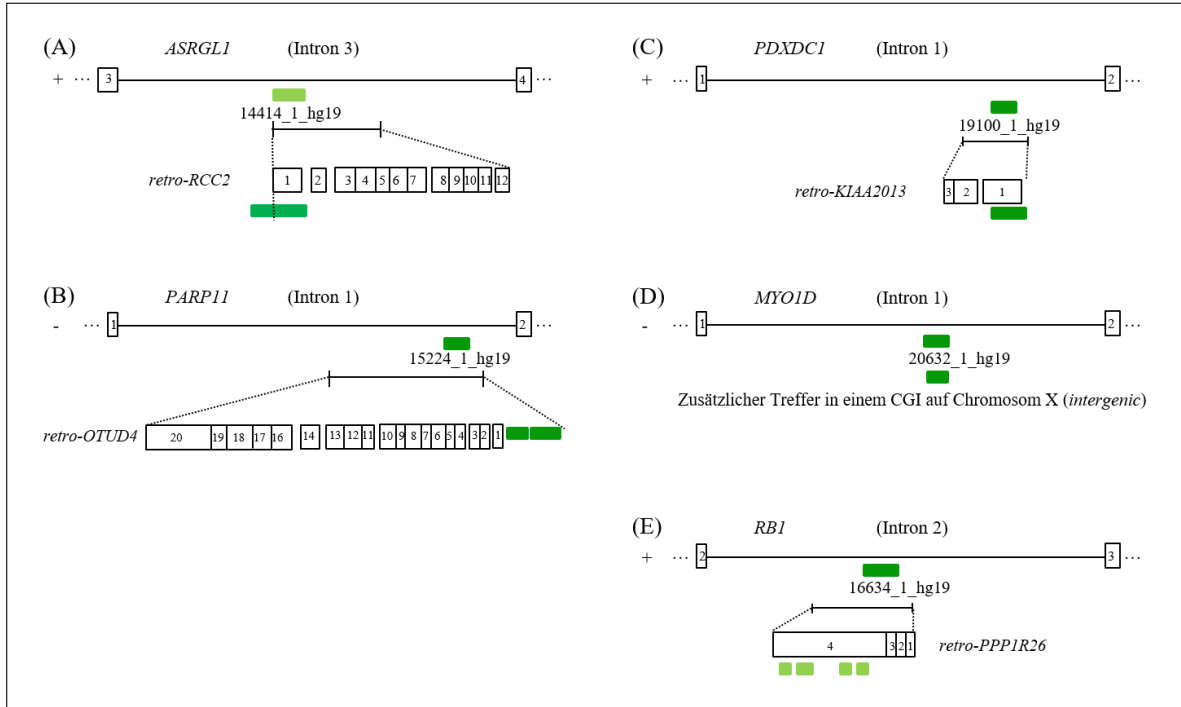


Abbildung 3.5: Schematische Darstellung der Introns, die ein CGI mit einem mittleren Methylierungsgrad beinhalten

Die Abbildung zeigt die Lage der CGIs (in grün dargestellt; Länge < 300bp: hellgrün) innerhalb der Gene und ihren vermutlichen Ursprung. Dargestellt ist das Intron, in dem das CGI liegt, und die flankierenden Exons (ohne Skalierung von Größen). „+“ und „-“ geben den Strang an, auf dem das exprimierte Gen liegt. (A) Intron 3 des *ASRGL1*-Gens enthält das CGI 14414_1_hg19 und eine Retrokopie des Gens *RCC2* (*retro-RCC2*). (B) Intron 1 des *PARP11*-Gens enthält das CGI 15224_1_hg19 und die Retrokopie *retro-OTUD4*. (C) Intron 1 des *PDXDC1*-Gens mit dem CGI 19100_1_hg19 und der Retrokopie *retro-KIAA2013*. (D) Intron 1 des *MYO1D*-Gens mit dem CGI 20632_1_hg19 und dem vermutlichen Ursprung, einem CGI auf dem X-Chromosom. (E) Intron 2 des *RB1*-Gens mit dem enthaltenen CGI 16634_1_hg19 (CpG85) und der Retrokopie *retro-PPP1R26*. Diese Abbildung wurde in abgeänderter Form bereits in Rademacher u. a. (2014) veröffentlicht.

Die Sequenzübereinstimmung des CGIs 20632_1_hg19 (*MYO1D*) liegt nicht in einem Gen, sondern in einem CGI auf dem X-Chromosom. Der Grad der Methylierung dieses ancestralen CGIs liegt in Monozyten bei 40%. Laut dem *UCSC Retroposed Gene Track* ist nur eine dieser vier Retrokopien (*retro-KIAA2013*) stark exprimiert (*strongly expressed*). Die anderen drei Retrokopien (*retro-RCC2*, *retro-OTUD4* und *retro-PPP1R26*) sind hingegen nur schwach exprimiert (*weakly expressed*).

3.2 Genomweite Analysen von Retrokopie-assoziierten CGIs

3.2.1 Identifikation von Retrokopie-assoziierten CGIs und deren Klassifizierung

Um herauszufinden, welche CGIs mit Retrokopien überlappen, wurden im ersten Schritt die genomischen Positionen von 27.537 humanen CGIs mit denen von 13.173 humanen Retrokopien verglichen (siehe Abschnitt 2.2.5). Hierbei zeigte sich, dass 10% (1.319) der Retrokopien mit CGIs überlappen (im weiteren Verlauf der Arbeit werden Retrokopie-assoziierte CGIs als R-CGIs bezeichnet). Die Abbildung 3.6 zeigt schematisch die Analyseschritte der genomweiten Analyse von R-CGIs.

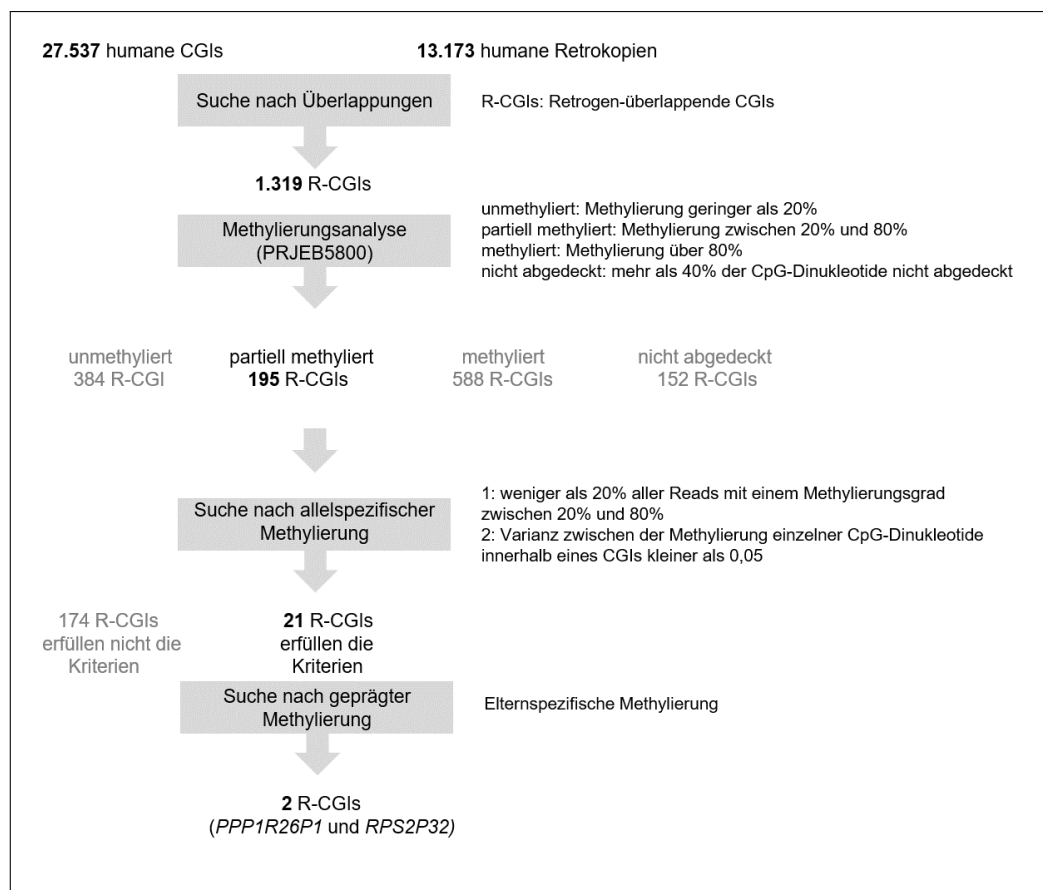


Abbildung 3.6: Flußdiagramm der genomweiten Analyse von Retrokopie-assoziierten CGIs

Die Abbildung gibt einen schematischen Überblick über die einzelnen Analyseschritte der genomweiten Analyse von Retrokopie-assoziierten CGIs inklusive der Filterkriterien. Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

Anschließend wurde die relative Lage dieser 1.319 Retrokopie-assoziierten CGIs zu Genen (RefSeq-Datensatz, siehe Abschnitt 2.1.4) untersucht. Eine Hälfte der R-CGIs liegt innerhalb von Genen (*intragenic*) und die andere Hälfte zwischen Genen (*intergenic*) bzw. in nicht-kodierenden Pseudogenen.

Im nächsten Schritt wurde die Expression dieser 1.319 R-CGIs genauer untersucht (Informationen über die Expression sind im Retrokopie-Datensatz vorhanden, siehe Abschnitt 2.1.3). Es zeigte sich, dass der Großteil der Retrokopien, die mit einem CGI überlappen, exprimiert werden (1.122; 85%). Nur ein kleiner Teil dieser Retrokopien (197; 15%), die mit CGIs überlappen, wurde innerhalb des UCSC Browsers als Pseudogen klassifiziert und werden somit nicht exprimiert.

Anschließend wurde die Häufigkeit eines R-CGIs im humanen Genom untersucht (siehe Tabelle 3.6).

Tabelle 3.6: Häufigkeit von Retrokopie-assoziierten CGIs

Vorkommen eines R-CGI(#)	Anzestrale Gene (#)	CGIs (#)
1	541	541
2	134	268
3	41	123
4	26	104
5	8	40
6	7	42
7	7	49
8	2	16
9	4	36
10	1	10
11	2	22
12	0	0
13	3	39
14	0	0
15	0	0
16	1	16
nicht klassifizierbar	<i>n/a</i>	13
Summe	778	1319

Die Tabelle gibt einen Überblick über die Häufigkeit von unabhängigen Retrotranspositionen der 1.319 R-CGIs. Angegeben ist neben der Anzahl der unabhängigen Transpositionen im Humangenom und der Anzahl von ancestralen Genen auch die absolute Anzahl von CGIs.

n/a: kein Gen-Name im RefSeq-Eintrag vorhanden. Diese Tabelle wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

Es gibt 134 anzestrale Gene, die mit einem CGI überlappen und genau zweimal im Humangenom vorkommen. Insgesamt wären dies dann 268 R-CGIs ($2 * 134$). Ungefähr die Hälfte aller R-CGIs ($n=541$) kommen nur einmal im humanen Genom vor. Mit zunehmender Anzahl eines R-CGIs nimmt die Anzahl an anzestralen Genen deutlich ab.

3.2.2 Methylierungsanalysen von Retrokopie-assoziierten CGIs

Die *in silico* Methylierungsanalysen der 1.319 R-CGIs wurden, wie auch im ersten Teil dieser Arbeit, an dem bereits publizierten Monozyten MethyloM 1 durchgeführt. Abbildung 3.7 zeigt die Ergebnisse der Methylierungsanalysen aller R-CGIs in Abhängigkeit der Anzahl von unabhängigen Retrotranspositionen.

Es gibt kein anzestrales Gen, dessen R-CGIs alle einen bestimmten Methylierungsgrad aufweisen, sondern dieser kann sich zwischen den einzelnen R-CGIs eines anzestralen Gens unterscheiden.

Insgesamt konnte von 1.167 R-CGIs der Grad der Methylierung ermittelt werden. Bei 152 R-CGIs war dies nicht möglich, da über 40% der CpG-Dinukleotide in dem MethyloM-Datensatz nicht abgedeckt waren. Diese R-CGIs wurden für die weiteren Analysen innerhalb dieser Arbeit ausgeschlossen. Die meisten (972; 85%) der 1.167 R-CGIs sind entweder vollständig methyliert (588; 50%; $m > 80\%$) oder unmethyliert (384; 32%; $m < 20\%$). Nur 195 (17%) der R-CGIs weisen einen mittleren Methylierungsgrad ($20\% \geq m \leq 80\%$) auf.

Im nächsten Schritt (siehe Abbildung 3.6) wurde untersucht, ob die Methylierung abhängig von der Expression der Retrokopie oder von der relativen Lage des R-CGIs ist. Es konnten jedoch keine Unterschiede in der Methylierung festgestellt werden, weder zwischen R-CGIs, die zwischen Genen (*intergenic*) bzw. innerhalb von Genen (*intragenic*) liegen, noch zwischen R-CGIs mit einer exprimierten Retrokopie und R-CGIs mit einer nicht-exprimierten Retrokopie (siehe Tabellen A.15, A.16, A.17 und A.18 im Anhang Abschnitt A.4.5). Im nächsten Schritt wurde das Methylierungslevel der direkten Umgebung der 1.167 R-CGIs untersucht, für die qualitativ hochwertige MethyloMdaten vorhanden sind. Die direkte Umgebung wurde hierfür als 1000bp *upstream* und *downstream* der Retrokopie definiert, um herauszufinden, ob die Methylierung der Integrationsstelle einer Retrokopie Einfluss auf die Methylierung der Retrokopie hat. Faktoren, die zusätzliche Einflüsse auf die Methylierung haben können, jedoch außerhalb dieser Region liegen, werden in dieser Arbeit nicht berücksichtigt.

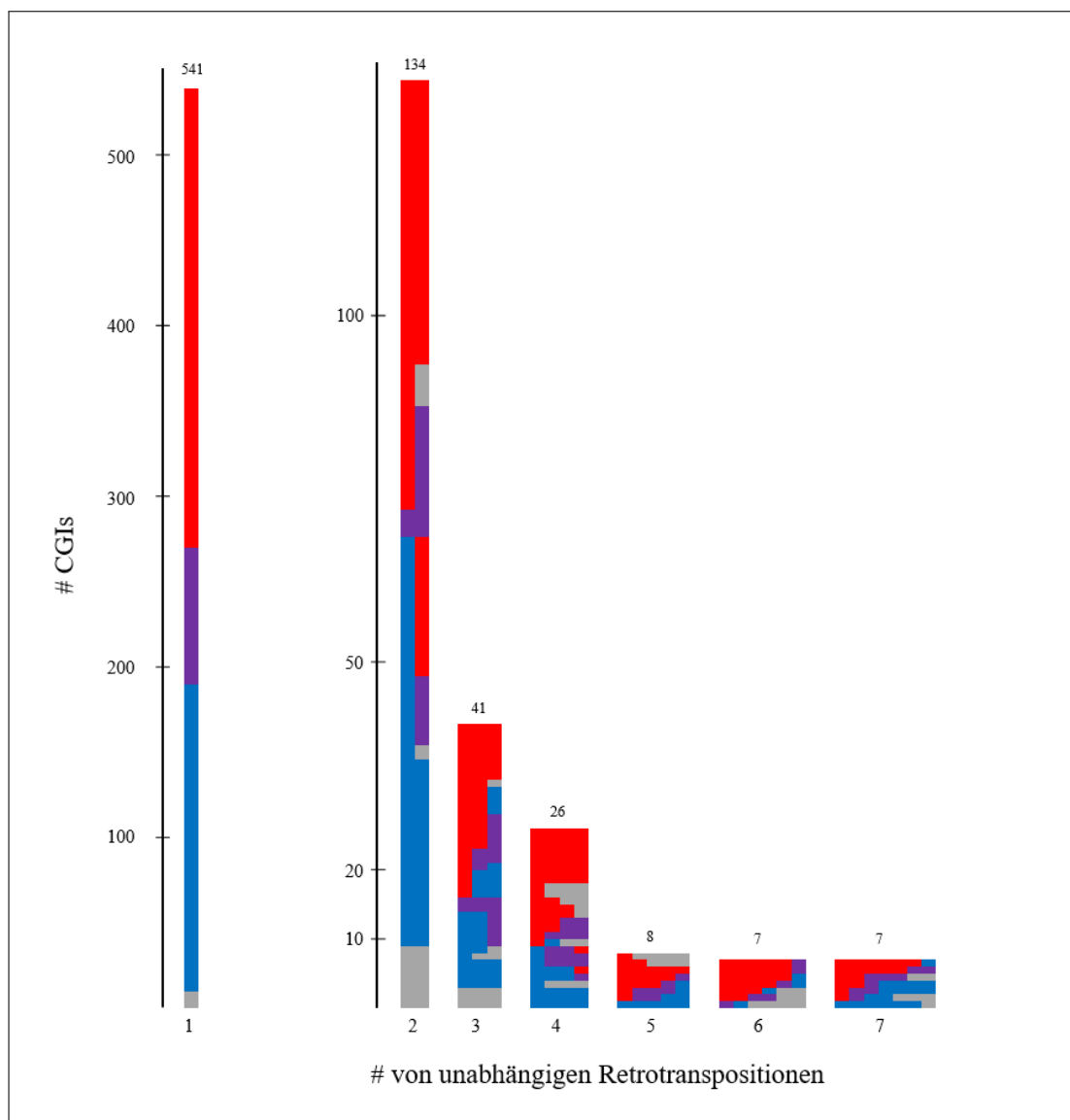


Abbildung 3.7: Grad der Methylierung der Retrokopie-assoziierten CGIs in Abhängigkeit der Anzahl ihrer Vorkommen

Die Abbildung zeigt den Methylierungsgrad der R-CGIs in Abhängigkeit von der Anzahl an Vorkommen der unabhängigen Retrotranspositionen. Jede einzelne Zelle steht für ein R-CGI. Die Farbe der Zelle gibt die Methylierung dieses CGIs an (rot: methylierte CGIs; blau: unmethylierte CGIs; lila: Methylierung zwischen 20% und 80%; grau: CGIs bei denen weniger als 40% der CpG-Dinukleotide abgedeckt sind). Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

3.2.2.1 Methylierungsanalysen der Umgebung von Retrokopie-assoziierten CGIs

Der Methylierungsgrad der Regionen 1000bp *upstream* und *downstream* der Retrokopie wurde analysiert und mit dem Methylierungsgrad des R-CGIs verglichen. Im ersten Schritt dieser Analyse wurde der Pearson Korrelationskoeffizient berechnet, um herauszufinden, ob es einen Zusammenhang zwischen der Methylierung des R-CGIs und der Umgebung der Retrokopie gibt. Das Ergebnis der Korrelationsanalyse ($r=0,78$) weist einen starken positiven Zusammenhang auf. Aufgrund dieses Ergebnisses wurde eine Clusteranalyse mit dem *k-means* Algorithmus durchgeführt (siehe Abschnitt 2.2.8). Hierbei wurde eine Anzahl von vier Clustern ($k=4$) gewählt. Die Ergebnisse dieser Analyse sind in der Abbildung 3.8 visualisiert.

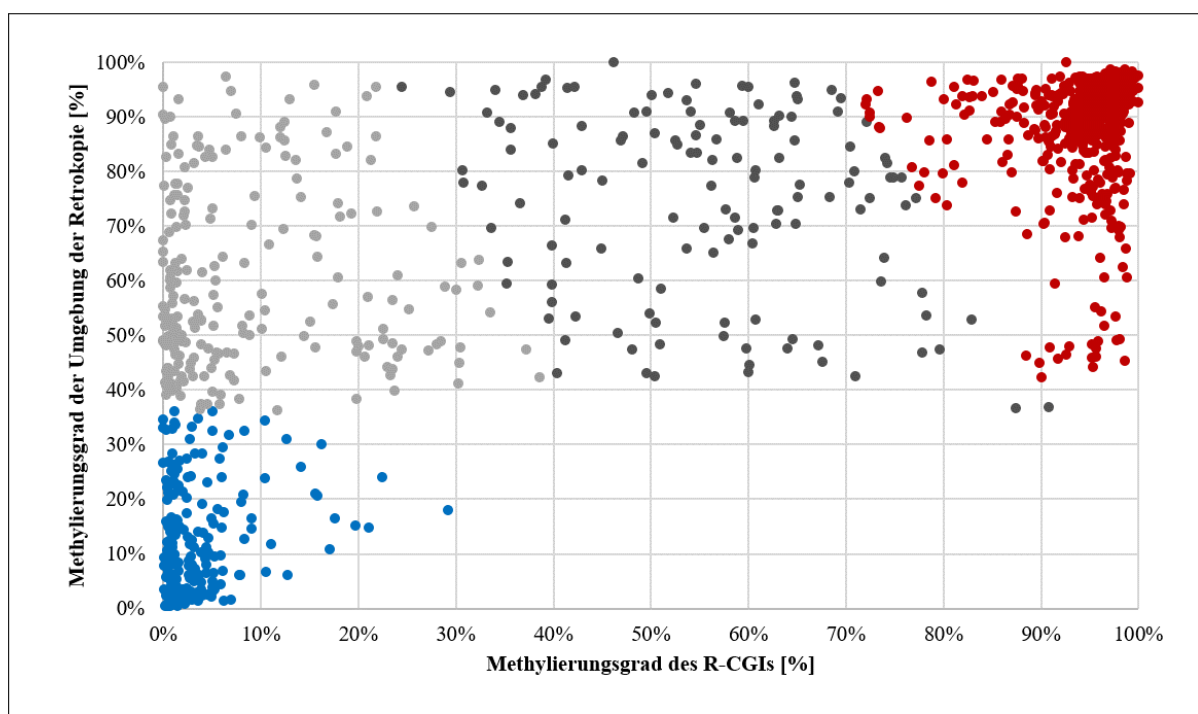


Abbildung 3.8: Clusteranalyse vom Methylierungsgrad der R-CGIs in Abhängigkeit vom Methylierungsgrad der Umgebung der Retrokopie

Die Abbildung zeigt den Grad der Methylierung der R-CGIs in Abhängigkeit des Methylierungsgrades der Umgebung der Retrokopie. Jeder einzelne Punkt repräsentiert den Methylierungsgrad eines R-CGIs (X-Achse) und den dazugehörigen Methylierungsgrad der Umgebung der Retrokopie (Y-Achse) in Prozent. Cluster 1: dunkelgrau, 136 Elemente (12%), Zentroid (56%/74%); Cluster 2: rot, 603 Elemente (52%), Zentroid (95%/89%); Cluster 3: blau, 204 Elemente (17%), Zentroid (4%/13%); Cluster 4: hellgrau, 224 Elemente (19%), Zentroid (9%/60%). Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

Mehr als die Hälfte (52%) der R-CGIs sind methyliert und auch die Umgebung der Retrokopie ist methyliert. Im Gegensatz zu methylierten R-CGIs, welche nicht in einer unmethylierten Umgebung auftreten, können unmethylierte R-CGIs sowohl in einer methylierten (19%) als auch unmethylierten (12%) Umgebung gefunden werden. Nur 195 R-CGIs haben einen Methylierungsgrad zwischen 20% und 80%. Diese R-CGIs findet man nur in Regionen mit einem Methylierungslevel zwischen 40% und 80%.

Zusätzlich wurde ein möglicher Einfluss von repetitiven Sequenzen auf die Methylierung der Umgebung von R-CGIs untersucht. In den meisten Fällen (n=839, 72%) waren sowohl in der *upstream* als auch in der *downstream* Region der Retrokopie repetitive Sequenzen vorhanden. Die Abwesenheit von repetitiven Sequenzen in beiden Regionen (n=47), ausschließlich in der *upstream* Region (n=137) oder ausschließlich in der *downstream* Region (n=144) ändert das Ergebnis der Clusteranalyse nicht (siehe Abbildungen A.17, A.18, A.19 und A.20 im Anhang A.8).

3.2.2.2 Vergleich der Methylierung des ancestralen CGIs mit der Methylierung des R-CGIs

Fast alle ancestralen Gene (96%) der R-CGIs besitzen mindestens ein CGI. Um herauszufinden, ob ein R-CGI im Laufe der Evolution neu entstanden ist oder aber mit einem CGI des ancestralen Gens übereinstimmt, wurden die Sequenzen des R-CGIs mit denen von CGIs des ancestralen Gens verglichen. Hierbei kam heraus, dass nur 372 der R-CGIs (32%) eine Sequenzidentität von mehr als 60% zu einem ancestralen CGI aufweisen. Dieses Ergebnis lässt darauf schließen, dass die Mehrheit der CGIs erst nach der Retrotransposition entstanden ist. Schaut man sich die Methylierung der 372 R-CGIs mit einer hohen Sequenzähnlichkeit zum ancestralen CGI an, so ist in den meisten Fällen (310) das ancestrale CGI unmethyliert, wobei jedoch die meisten der R-CGIs methyliert sind (262). Dieses Ergebnis lässt auf einen Zugewinn von Methylierung nach der Retrotransposition schließen.

3.2.2.3 Suche nach allelspezifischer Methylierung

Um herauszufinden, ob eine allelspezifische Methylierung bei einem R-CGI vorliegt, wurden die 195 R-CGIs mit einem Methylierungsgrad zwischen 20% und 80% genauer analysiert.

Hierbei wurde, wie auch in Abschnitt 3.1.2, die Methylierung einzelner Reads eines CGIs genauer untersucht. In einer Vorauswahl wurden R-CGIs mit einem methylierten und einem unmethylierten Ende ausgeschlossen. Auch hier wurden die gleichen Kriterien wie im ersten Teil dieser Arbeit benutzt (Anteil partiell methylierter Reads $\leq 20\%$ und VAR $\leq 0,05$), um neue differentiell methylierte CGIs zu identifizieren. Mit diesem Ansatz wurden 17 neue CGIs (siehe Tabelle 3.7) identifiziert, wobei 15 CGIs beide Kriterien erfüllen und zwei CGIs jeweils nur ein Kriterium. Bei diesen 17 R-CGIs könnte das mittlere Methylierungsniveau Ergebnis einer allelspezifischen Methylierung sein. Neben diesen 17 CGIs (Screenshots einzelner Reads dieser 17 CGIs sind auf der Daten-CD zu finden), wurden ebenfalls die vier human/nicht-murin intronischen CGIs (14414_1_hg19, 15224_1_hg19, 16634_1_hg19 und 19100_1_hg19) aus Abschnitt 3.1.2 identifiziert, die mit einer Retrokopie überlappen (Anmerkung: CGI 20632_1_hg19 aus Abschnitt 3.1.2 überlappt nicht mit einer Retrokopie). Diese vier CGIs werden in den folgenden Ergebnissen nicht mehr berücksichtigt.

Die Methylierung dieser 17 R-CGIs wurde anschließend auch in MethyloM 2, in den 16 bereits publizierten MethyloMen (aus Ziller u. a. (2013); unterschiedliche Gewebetypen siehe auch Abschnitt 3.1.2) und zwei Oozyten-MethyloMen (publiziert von Okae u. a. (2014)) untersucht (siehe Tabellen A.20, A.21 und A.22 im Anhang). Alle 17 untersuchten R-CGIs sind in Spermien unmethyliert. In Oozyten sind die untersuchten R-CGIs entweder vollständig methyliert (vier R-CGIs: 9224_1_hg19, 9261_1_hg19, 15521_1_hg19 und 24982_1_hg19) oder aber vollständig unmethyliert (13 R-CGIs). Eine Methylierung zwischen 20% und 80% zeigten sieben der 17 R-CGIs (2391_1_hg19, 9224_1_hg19, 9261_1_hg19, 9377_1_hg19, 17031_1_hg19, 23548_1_hg19 und 24982_1_hg19) in fast allen Gewebetypen. Drei R-CGIs (13250_1_hg19, 16458_1_hg19 und 20403_1_hg19) wiesen in fast allen Geweben keine Methylierung auf und sechs R-CGIs (9471_1_hg19, 13085_1_hg19, 15400_1_hg19, 15512_1_hg19, 16448_1_hg19) sind in mindestens drei Geweben methyliert (in den anderen Gewebetypen liegt eine Methylierung zwischen 20% und 80% vor). Das R-CGI 6141_1_hg19 zeigte weder in Spermien noch in den untersuchten fetalen Geweben eine Methylierung. In adulten Geweben weist dieses CGI ein mittleres Methylierungsniveau auf. Eine ausführliche Methylierungsanalyse dieser Gewebetypen war nicht möglich, da keine Readdaten zur Verfügung standen.

Tabelle 3.7: Methylierungsanalyse einzelner Reads und CpG-Dinukleotide der Retrokopie-assoziierten CGIs mit einem mittleren Methylierungsgrad (Methylohm 1)

CGI ID	Genomische Lage	Anzestrales Gen	\emptyset Methylierung (%)	\emptyset Abdeckung	# Reads	unmeth. Reads (#)	(%)	meth. Reads (#)	(%)	part. meth. Reads (#)	(%)	VAR	SD
2391_1_hg19	chr1:240656253-240656720	<i>GREM1</i>	56	12	83	29	35	44	53	10	12	0,04	0,20
6141_1_hg19	chr4:144833114-144833512	<i>SAV1</i>	49	18	111	67	60	39	35	5	5	0,04	0,20
9224_1_hg19	chr7:16890768-16891087	<i>ARHGAP20</i>	58	13	62	24	39	27	44	11	18	0,06	0,23
9261_1_hg19	chr7:23530434-23530690	<i>RPS2</i>	72	15	59	16	27	40	68	3	5	0,01	0,11
9377_1_hg19	chr7:36010997-36011407	<i>PPPIR14B</i>	39	7	48	25	52	15	31	8	17	0,04	0,21
9473_1_hg19	chr7:52341468-52342266	<i>CCDC115</i>	79	12	108	19	18	74	69	15	14	0,03	0,17
13085_1_hg19	chr10:66813635-66814061	<i>NEK4</i>	76	8	58	9	16	40	69	9	16	0,02	0,16
13250_1_hg19	chr10:91596974-91597792	<i>MARK2</i>	22	10	109	83	76	23	21	3	3	0,01	0,07
15400_1_hg19	chr12:31405184-31405545	<i>RPL13AP5</i>	69	16	97	26	27	61	63	10	10	0,02	0,15
15512_1_hg19	chr12:49782965-49783193	<i>FGFR1OP2</i>	54	18	59	22	37	30	51	7	12	0,02	0,15
16448_1_hg19	chr13:21893156-21893605	<i>GRK6</i>	79	13	88	9	10	67	76	12	14	0,02	0,12
16458_1_hg19	chr13:23412207-23412623	<i>IPMK</i>	31	11	66	41	62	17	26	8	12	0,02	0,15
17031_1_hg19	chr14:21191657-21191860	<i>XPO6</i>	42	7	30	18	60	11	37	1	3	0,02	0,16
20403_1_hg19	chr17:15686218-15686474	<i>IL6ST</i>	20	13	49	41	84	7	14	2	4	0,01	0,10
23548_1_hg19	chr19:21860792-21861016	<i>MTDH</i>	59	14	52	14	27	27	52	11	21	0,03	0,17
24982_1_hg19	chr20:30135076-30135292	<i>MCTS1</i>	55	20	64	25	39	37	58	2	3	0,01	0,10
26859_1_hg19	chrX:37026348-37026706	<i>FAM47A</i>	77	8	44	6	14	34	77	4	9	0,02	0,15

Diese Tabelle zeigt die Ergebnisse der Methylierungsanalyse einzelner Reads und CpG-Dinukleotiden der 17 potentiell differentiell methylierten R-CGIs für das Methylohm 1. Angegeben sind die Namen der CGIs (CGI ID), sowie Gen und Retrokopie, Chromosom, Länge, Durchschnittsmethylierung, Durchschnittsabdeckung und die Anzahl analysierter Reads. Zusätzlich sind die einzelnen Reads in drei Klassen unterteilt: Unmethylierte ($m < 20\%$), methylierte ($m > 80\%$) und partiell methylierte Reads ($20\% \geq m \leq 80\%$). Die letzten beiden Spalten zeigen die Ergebnisse der Methylierungsanalyse der CpG-Dinukleotide. Varianz (VAR) und Standardabweichung (SD) aller einzelner CpGs eines CGIs wurden berechnet. Ein großer Teil der R-CGIs (15 von 17) erfüllten beide Kriterien (Anteil partiell methylierter Reads $\leq 20\%$ und VAR $\leq 0,05$), zwei R-CGIs erfüllen nur ein Kriterium. Diese Tabelle wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

3.2.2.4 Methylierungsanalysen mittels Roche/454 GS Junior

Im nächsten Schritt wurden Amplikons für die 17 R-CGIs generiert, um diese mittels hochquantitativer Hochdurchsatz-Bisulfitssequenzierung auf dem Roche/454 GS Junior zu sequenzieren. Ziel dieser Untersuchung war es herauszufinden, ob die Methylierung dieser 17 R-CGIs allelspezifisch ist. Es wurde DNA von 21 Monozyten (unverwandte, männliche Spender) auf unterschiedliche SNPs in der Region der 17 R-CGIs untersucht. Insgesamt konnten für zwölf R-CGIs (2391_1_hg19, 6141_1_hg19, 9224_1_hg19, 9261_1_hg19, 9473_1_hg19, 13250_1_hg19, 15400_1_hg19, 16448_1_hg19, 17031_1_hg19, 20403_1_hg19 und 23548_1_hg19) heterozygote Monozyten-Proben für einen untersuchten SNP identifiziert werden (siehe Tabellen A.6, A.7, A.8 und A.9 im Anhang).

Tabelle 3.8: Methylierungsanalyse der zwölf R-CGIs mittels Roche/454 GS Junior

CGI ID	Probe	Allel a		CpGs (#)	Reads (#)	Allel b		CpGs (#)	Reads (#)
		Ø	Meth. (%)			Ø	Meth. (%)		
2391_1_hg19	R3	59%		13	714	G	60%	213	478
	R9	57%		13	607	G	60%	135	458
6141_1_hg19	R7	28%		41	1452	A	47%	393	1057
	R13	23%		41	1127	A	39%	366	759
9224_1_hg19	R6	32%		13	2179	G	8%	960	1150
	R7	23%		13	2800	G	26%	1226	1484
9261_1_hg19	R5	64%		12	897	G	87%	536	326
	R13	47%		12	1002	G	23%	428	542
9473_1_hg19	R8	77%		18	2491	A	86%	1393	929
	R10	79%		18	2542	A	86%	1369	1017
13250_1_hg19	R3	3%		19	1801	G	4%	677	934
	R5	4%		19	3060	G	6%	1338	1499
15400_1_hg19	R4	70%		29	1840	G	62%	1012	810
	R14	78%		29	1482	G	82%	1063	411
16448_1_hg19	R3	61%		31	516	A	81%	240	272
	R5	94%		31	982	A	94%	406	566
16458_1_hg19	P2	22%		27	984	A	15%	302	509
17031_1_hg19	R4	29%		18	4187	G	19%	2038	2124
	R14	17%		18	4814	G	12%	2313	2469
20403_1_hg19	R1	3%		32	3397	T	2%	1503	1889
	R8	3%		32	2255	T	3%	1116	1135
23548_1_hg19	R2	25%		22	3633	G	38%	1910	1700
	R15	25%		22	3438	G	19%	1569	1851

Diese Tabelle zeigt die Ergebnisse der Methylierungsanalysen mittels hochquantitativer Hochdurchsatz-Bisulfitssequenzierung auf dem Roche/454 GS Junior der zwölf R-CGIs. Neben der ID des CGIs sind Monozyten-Probe, durchschnittliche Methylierung, Anzahl analysierter CpGs und die Gesamtzahl von Reads angegeben. Zusätzlich sind die Ergebnisse nach den Allelen aufgetrennt dargestellt (Allel a und Allel b). Angegeben sind für jedes Allel das Allel selbst, die durchschnittliche Methylierung und die Readzahlen. Diese Tabelle wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

Es wurden Amplikons für diese zwölf R-CGIs generiert und die Methylierung mittels Roche/454 GS Junior untersucht. Für elf dieser zwölf R-CGIs konnte die Methylierung von zwei informativen Proben untersucht werden, bei einem R-CGI von einer informativen Probe (16458_1_hg19). Tabelle 3.8 zeigt die Ergebnisse der Methylierungsanalysen mittels

Roche/454 GS Junior. Die Methylierungsplots der einzelnen Proben sind im Anhang A.4.7 Abbildungen A.13, A.14, A.15 und A.16 zu finden.

Im Gegensatz zu dem Ergebnis der *in silico* Analyse scheinen zwei der untersuchten R-CGIs (13250_1_hg19 und 20403_1_hg19) in Monozyten vollständig unmethyliert zu sein. Dieses Ergebnis konnte auch schon durch die Methylierungsanalysen in bereits veröffentlichten Methylogen gezeigt werden (siehe Tabellen A.21 und A.22). Bei insgesamt fünf R-CGIs (6141_1_hg19, 9261_1_hg19, 9473_1_hg19, 15400_1_hg19 und 23548_1_hg19) konnten Unterschiede in der Methylierung der beiden Allele von mindestens 10% festgestellt werden. Zwei dieser fünf R-CGIs (6141_1_hg19 und 9473_1_hg19) zeigen in beiden untersuchten Proben jeweils, dass „Allel a“ deutlich stärker methyliert ist als „Allel b“. Diese Ergebnisse können auf einen *cis*-agierenden SNP hindeuten. Bei den drei übrigen R-CGIs (9261_1_hg19, 15400_1_hg19 und 23548_1_hg19) konnten ebenfalls Methylierungsunterschiede von mehr als 10% zwischen den Allelen beobachtet werden. Im Gegensatz zu den zwei R-CGIs, bei denen jeweils ein Allel immer eine höhere Methylierung als das andere Allel aufweist, finden sich bei diesen drei R-CGIs unterschiedliche Methylierungsgrade eines Allels in den untersuchten Proben. Sowohl „Allel a“ als auch „Allel b“ kommen jeweils einmal mehr methyliert bzw. weniger methyliert vor. Diese Ergebnisse können ein Hinweis auf genomische Prägung (siehe Abschnitt 1.4) sein. Das R-CGI 9261_1_hg19 zeigt mit mehr als 40% Unterschied die größten allelischen Methylierungsunterschiede aller untersuchten R-CGIs. Basierend auf diesen Ergebnissen wurde eine genauere Analyse dieses R-CGI angeschlossen.

3.2.3 Das R-CGI 9261_1_hg19

Das R-CGI 9261_1_hg19 liegt auf Chromosom 7 (siehe Abbildung 3.9 A) und überlappt in seiner genomischen Position mit der Retrokopie *retro-RPS2*. Im menschlichen Genom gibt es insgesamt 57 Vorkommen dieser Retrokopie, wobei es nur sechsmal in Verbindung mit einem CGI auftritt. Alle fünf weiteren Vorkommen des *retro-RPS2* mit einem CGI sind in Monozyten (Methylom 1) methyliert. Einzige Ausnahme ist das R-CGI 9261_1_hg19, das in allen untersuchten Geweben (außer in Spermien, in denen das R-CGI unmethyliert ist und Oozyten, in denen es methyliert ist) ein mittleres Methylierungsniveau aufweist. Das anzestrale *RPS2*-Gen liegt auf Chromosom 16 und überlappt mit einem großen CGI (ca. 1200bp). Sowohl CGI (9261_1_hg19) als auch Retrokopie (*retro-RPS2*) liegen innerhalb des *RPS2P32*-Gens (RefSeq ID: NR_026676.1). Laut UCSC Browser handelt es sich

bei *RPS2P32* um ein Pseudogen (*Homo sapiens ribosomal protein S2 pseudogene 32*). Abbildung 3.9 (B) zeigt die benachbarten Gene von *RPS2P32*. Ca. 20kb telomerisch liegt das *IGF2BP3*-Gen und etwa 23kb centromerisch das *TRA2A*-Gen. Das *TRA2A*-Gen kodiert ein RNA-bindendes Protein, das in die Regulation des Spleißens (prä-mRNA) involviert ist. *IGF2BP3* kodiert ein Protein, das an die mRNA des *IGF2*-Gens bindet und dessen Translation unterdrücken kann (Monk u. a., 2002).

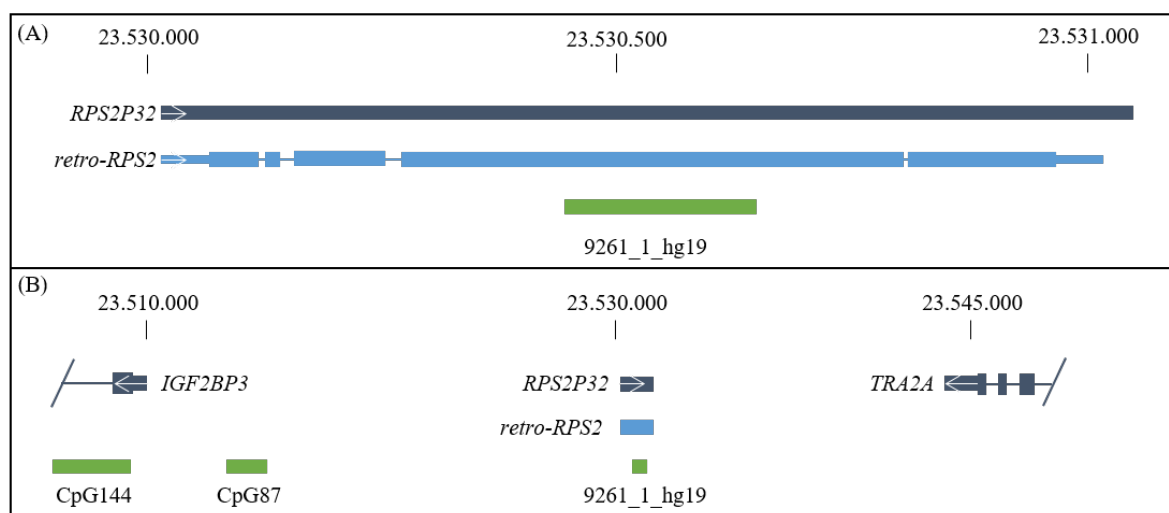


Abbildung 3.9: Schematische Darstellung der Lage des R-CGIs 9261_1_hg19 im humanen Genom

Die Abbildung zeigt die genomische Lage des R-CGIs 9261_1_hg19 auf Chromosom 7 des humanen Genoms. RefSeq-Gene sind in dunkelblau, Überlappungen mit einer Retrokopie in hellblau und CGIs in grün dargestellt. Weiße Pfeile zeigen die Richtung der Transkription an. Die Zahlen spiegeln die Position auf Chromosom 7 wider (hg19). (A) zeigt das *RPS2P32*-Gen, das R-CGI (9261_1_hg19) und die mit dem ancestralen *RPS2*-Gen überlappenden Basen. In (B) wird die relative Lage im Humangenom deutlich. Ein größerer Ausschnitt der Region mit den benachbarten Genen (telomerisch: *IGF2BP3*-Gen, centromerisch: *TRA2A*-Gen) ist zu sehen. Die Abbildung ist nicht maßstabsgetreu gezeichnet und wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

3.2.3.1 Chromatin-Daten zu R-CGI 9261_1_hg19

Mit Hilfe von Daten aus dem Deutschen Epigenom Projekt (DEEP), an dem das Institut für Humangenetik beteiligt ist, konnte die Region um das R-CGI 9261_1_hg19 im Hinblick auf Histonmodifikation genauer untersucht werden. Wie in Abbildung 3.10 dargestellt, sind im Bereich des *RPS2P32*-Gens sowohl Promotor (H3K4me3 und H3K27ac) als auch Enhancer (H3K27ac und H3K4me1) aktiv. Die Transkription von *RPS2P32* ist somit aktiviert, was zusätzlich an der Histonmodifikation H3K36me3 deutlich wird. Desweiteren ist die

Zugänglichkeit der chromosomalen DNA in dieser Region gegeben, da das Heterochromatin zu Euchromatin entpackt wurde (*chromatin accessibility*, CA).

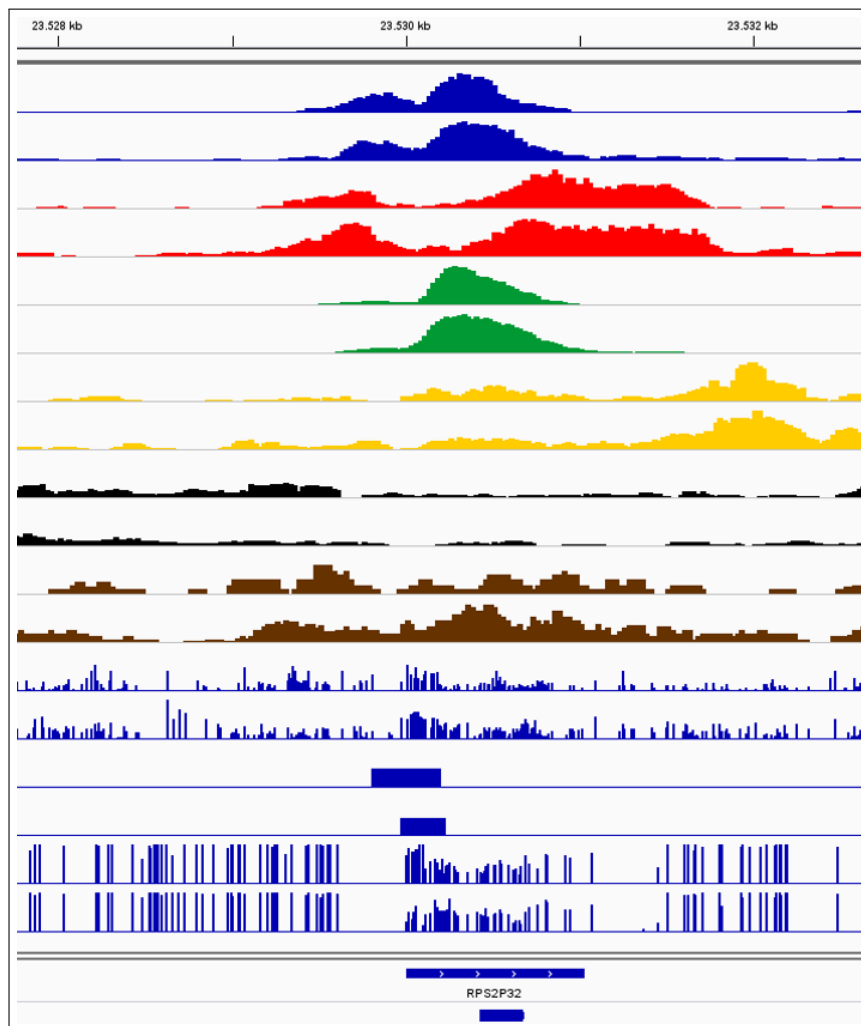


Abbildung 3.10: Histonmodifikationen des R-CGIs 9261_1_hg19

Die Abbildung zeigt einen Screenshot des IGV-Browsers von einem Abschnitt auf Chromosom 7 (chr7:23.527.800-23.532.500) der das *RPS2P32*-Gen und das R-CGI 9261_1_hg19 beinhaltet. Dargestellt sind für zwei Proben von unterschiedlichen Spendern (43_Hm04_BIMo_Ct, 43_Hm05_BIMo_Ct) die Histonmodifikation in Monozyten. Die folgenden Histonmodifikationen sind in der Abbildung dargestellt: H3K27ac (blau, oberen 2 Kurven, eine Kurve jeweils für einen Spender), H3K4me1 (rot), H3K4me3 (grün), H3K36me3 (gelb), H3K27me3 (schwarz) und H3K9me3 (braun). Die unteren sechs Graphen (jeweils zwei gehören zu einer Modifikation) in blau stehen für die Chromatin-Zugänglichkeit (*chromatin accessibility*, CA), einmal als Rohdaten und einmal als Peak und für die Methylierung. Zusätzlich sind noch das *RPS2P32*-Gen inklusive der Transkriptionsrichtung und das CGI 9261_1_hg19 in blau am unteren Rand dargestellt. Die Abbildung wurde freundlicherweise von Prof. Dr. Bernhard Horsthemke zur Verfügung gestellt.

3.2.3.2 Methylierungsanalyse des R-CGIs 9261_1_hg19 in informativen Eltern-Kind-Trios (Blut)

Um herauszufinden, ob die allelspezifische Methylierung des R-CGIs 9261_1_hg19 elternspezifischer Herkunft ist, wurde die Methylierung in drei informativen Eltern-Kind-Trios für den ausgewählten SNP (rs10228640) untersucht. Hierfür sollte das Kind heterozygot für den SNP sein und mindestens ein Elternteil homozygot. Für die Kinder wurden Amplikons generiert und auf dem Roche/454 GS Junior sequenziert (siehe Abbildung 3.11).

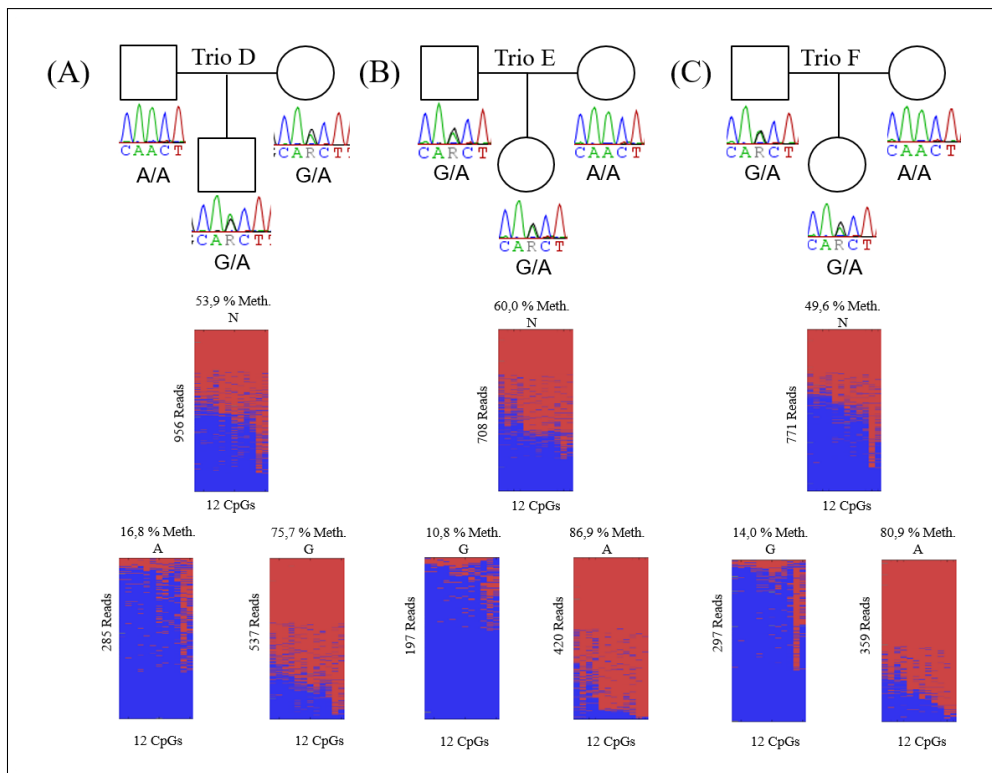


Abbildung 3.11: Methylierungsanalyse des R-CGIs 9261_1_hg19 in Blut von drei heterozygoten Individuen

Die Abbildung zeigt für jedes untersuchte Individuum einen Stammbaum inklusive der Elektropherogramme des untersuchten SNPs (rs10228640) und den Genotyp. Die Methylierungsplots unter dem jeweiligen Individuum zeigen die Methylierung aller Reads eines Amplikons. Unter diesem befinden sich die Plots aufgetrennt nach Allelen, wobei jeweils das paternal vererbte Allel links und das maternal vererbte Allel rechts angeordnet ist (analog zum Stammbaum). Über jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben. Links steht die Anzahl an Reads und unter dem Plot die Anzahl analysierter CpGs. Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. In allen drei untersuchten Trios (Trio D-F (A)-(C)) wurde jeweils das unmethylierte Allel paternal und das methylierte Allel maternal vererbt. Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

Alle drei untersuchten Kinder weisen ein mittleres Methylierungsniveau (zwischen 49% und 60%) auf. Wie auch schon bei den vorherigen Analysen in Monozyten beobachtet werden konnte, ist immer ein Allel deutlich stärker methyliert als das andere Allel. In Monozyten lagen die Unterschiede in der Methylierung der getrennten Allele bei etwas mehr als 40%. In Blut konnten sogar Methylierungsunterschiede von über 70% zwischen den beiden Allelen beobachtet werden. Dem Kind von Trio D (Abbildung 3.11 (A)) wurde das unmethylierte Allel (A) paternal und das methylierte Allel (G) maternal vererbt. Während den beiden Kinder der Trios E und F (Abbildung 3.11 (B) und (C)) jeweils das unmethylierte Allel (G) paternal und das methylierte Allel (A) maternal vererbt wurde. Alle drei Kinder haben somit das methylierte Allel maternal und das unmethylierte Allel paternal ererbt. Dieses Ergebnis schließt einen *cis*-agierenden Effekt der DNA-Methylierung aus und deutet auf eine elterliche Prägung der Methylierung. Desweiteren wurde die Methylierung von *RPS2P32* in Plazenta unterschiedlicher Trimester untersucht (siehe Abbildung 3.12). Auch in diesem Gewebe konnte ein Methylierungsgrad um die 50% festgestellt werden. Leider waren alle drei Proben homozygot für den untersuchten SNP, so dass eine Segregationsanalyse nicht möglich war.

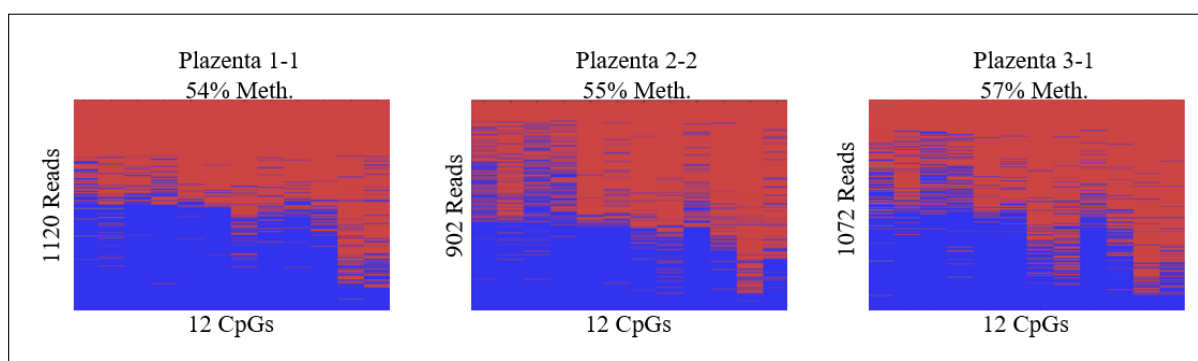


Abbildung 3.12: Methylierungsanalyse des R-CGIs 9261_1_hg19 in drei Plazenta-Proben

Die Abbildung zeigt die Methylierungsplots für drei verschiedene Plazenta-Proben. Über jedem Plot steht der Probenname, wobei die erste Zahl das Trimester der Probe und die zweite die interne Probennummer angibt und der durchschnittliche Methylierungsgrad aller Reads über alle CpGs. Links steht die Anzahl an Reads, unter dem Plot die Anzahl analysierter CpGs. Methylierte CpGs sind in rot dargestellt, unmethylierte CpGs in blau.

3.2.3.3 Methylierungsanalyse des R-CGIs 9261_1_hg19 in Patienten mit einem Multilocus-Imprinting-Defekt (Blut)

Zusätzlich zu den drei Eltern-Kind-Trios wurde die Methylierung des R-CGIs 9261_1_hg19 noch in sechs Patienten mit einem bestätigten Multilocus-Imprinting-Defekt

3 ERGEBNISSE

(MLID) und vier Normalkontrollen (NC) untersucht. Wie in der Abbildung 3.13 visualisiert ist, zeigen vier (MLID 3, MLID 4, MLID 5 und MLID 6) der sechs MLID-Patienten eine Hypomethylierung des R-CGIs 9261_1_hg19 auf. Bei den übrigen zwei MLID-Patienten (MLID 1 und MLID 2) wurde ein mittlerer Methylierungsgrad um die 50% festgestellt, genau wie bei den vier Normalkontrollen (NC 1, NC 2, NC 3 und NC 4).

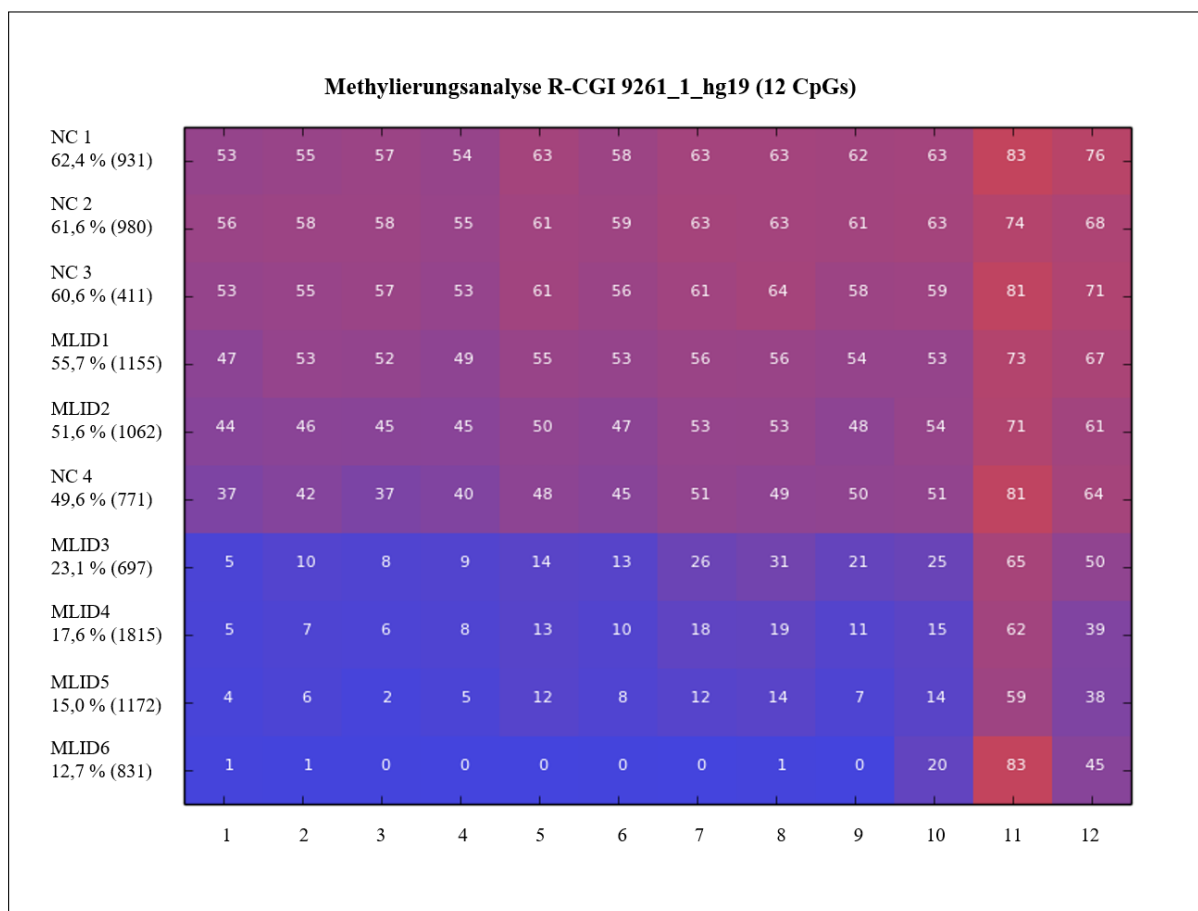


Abbildung 3.13: Methylierungsanalyse des R-CGIs 9261_1_hg19 von sechs Patienten mit einem Multilocus-*Imprinting*-Defekt und vier Normalkontrollen

Diese Abbildung zeigt die Ergebnisse der Methylierungsanalyse mittels Roche/454 GS Junior bei sechs Patienten mit einem bestätigten Multilocus-*Imprinting*-Defekt (MLID) und vier Normalkontrollen (NC). Pro analysiertem Individuum gibt es eine Zeile, die die Methylierung der einzelnen CpG-Dinukleotide zeigt. Unter dem Probenamen sind die durchschnittliche Methylierung der 12 untersuchten CpGs und die Anzahl der untersuchten Reads in Klammern angegeben. Die Farbe spiegelt den Grad der Methylierung jedes einzelnen CpGs wider, von blau unmethyliert nach rot methyliert. Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

3.2.3.4 Expressionsanalysen des R-CGIs 9261_1_hg19

Es wurde eine Expressionsanalyse an Blut von einem der drei untersuchten Kinder (Trio F) aus Abschnitt 3.2.3.2 (siehe Abbildung 3.11) durchgeführt, um herauszufinden, ob *RPS2P32* monoallelisch oder biallelisch exprimiert wird. Für die Individuen von Trio D und Trio E stand leider kein Material für eine Expressionsanalyse zur Verfügung. In dem untersuchten Kind wird das R-CGI 9261_1_hg19 biallelisch exprimiert. Die Expressionsanalyse für Trio F wurde auf beiden Strängen durchgeführt. Zusätzlich zur Sangersequenzierung wurde noch eine Primer-Verlängerung mittels SNaPshot-Analyse und anschließend Genescan durchgeführt. *RPS2P32* wird auf beiden Strängen in beiden Richtungen exprimiert. Während auf dem *forward*-Strang beide Allele in etwa gleich stark exprimiert (Rate=1,18) werden, konnte auf dem *reverse*-Strang eine kleine Verschiebung der Expression (Rate=1,78) zugunsten des paternalen (unmethylierten) Alleles festgestellt werden (Rohdaten der SNaPshot Analyse siehe Anhang Abschnitt A.9 Tabelle A.25).

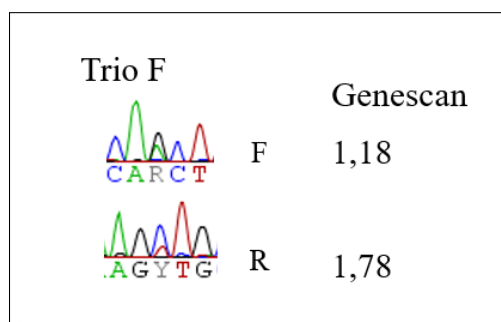


Abbildung 3.14: Expressionsanalyse des R-CGIs 9261_1_hg19 (Blut)

Diese Abbildung zeigt die Ergebnisse der Expressionsanalyse des R-CGIs 9261_1_hg19 für das Kind des Trios F (siehe Abbildung 3.11 Abschnitt 3.2.3.2). Die Analyse zeigt eine biallelische Expression des R-CGIs 9261_1_hg19 bei dem Kind auf beiden Strängen, wobei F für den *forward*-Strang und R für den *reverse*-Strang steht. Zusätzlich sind noch die Ergebnisse des Genescans (siehe Abschnitt 2.4.3.9) angegeben, die für den *forward*-Strang eine biallelische Expression bestätigen und für den *reverse*-Strang eine leichte Verschiebung zugunsten des paternalen Allels zeigen. Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

3.2.3.5 Analyse der benachbarten Gene *TRA2A* und *IGF2BP3*

Da die differentielle Methylierung des R-CGIs 9261_1_hg19 die Expression von *RPS2P32* in Blut nicht reguliert und keine anderen Gewebeproben zur Verfügung standen, wurden die Analysen auf die benachbarten Gene von *RPS2P32* ausgedehnt (siehe Abbildung 3.9). Es

konnte kein informativer SNP (zu geringe Allelfrequenzen) für *TRA2A* detektiert werden. Deshalb konnte dieses Gen nicht analysiert werden und die folgenden Analysen beschränken sich auf *IGF2BP3*. Das *IGF2BP3*-Gen besitzt zwei CGIs (CpG87 und CpG144; siehe Abbildung 3.9). Im ersten Schritt wurde die Methylierung beider CGIs von *IGF2BP3* in Plazenta untersucht, da *IGF2BP3* in Plazenta stark exprimiert ist. Wie in Abbildung 3.15 sichtbar, sind beide CGIs vollständig unmethyliert.

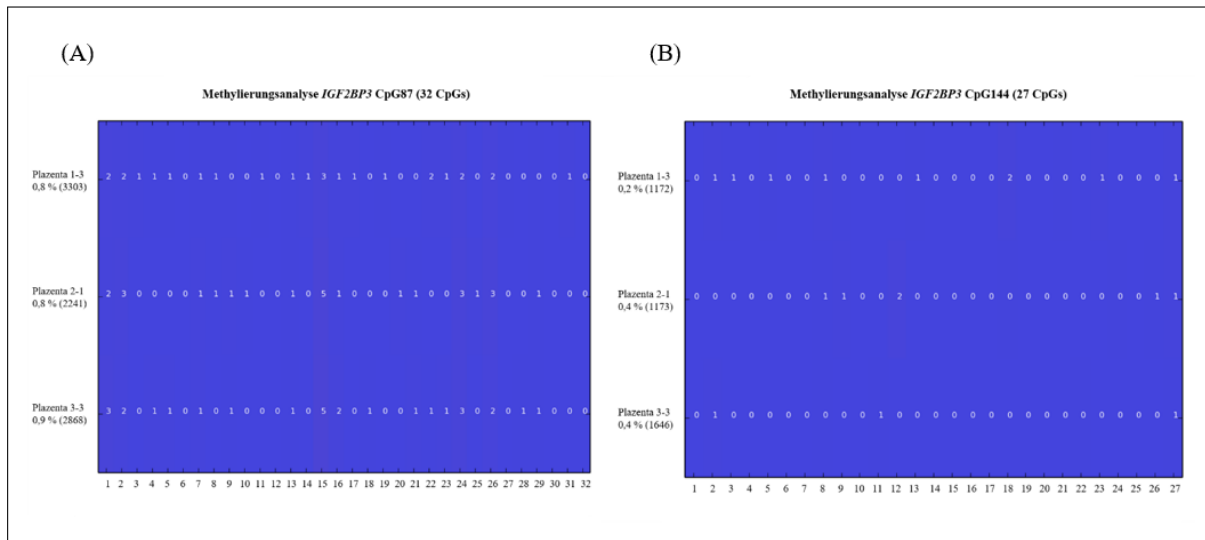


Abbildung 3.15: Methylierungsanalyse der beiden CGIs (CpG87 und CpG144) des *IGF2BP3*-Gens

Diese Abbildung zeigt die Ergebnisse der Methylierungsanalyse mittels Roche/454 GS Junior der beiden CGI des *IGF2BP3*-Gens in Plazenta aus den drei Trimestern. Pro analysierter Probe gibt es eine Zeile, die die Methylierung der einzelnen CpG-Dinukleotide zeigt. Links neben jeder Zeile ist neben dem Probenamen (erste Zahl gibt das Trimester an, die zweite die Probennummer), die durchschnittliche Methylierung der untersuchten CpGs und die Anzahl der untersuchten Reads in Klammern angegeben. Die Farbe spiegelt den Grad der Methylierung jedes einzelnen CpGs wider von blau unmethyliert nach rot methyliert. (A) zeigt die Ergebnisse des CpGs87 und (B) die Ergebnisse des CpGs144.

Im nächsten Schritt wurde die Expression des *IGF2BP3*-Gens (rs156414) analysiert, hierfür wurde sowohl RNA aus Blut (siehe Abbildung 3.16) als auch aus Plazenta (siehe Abbildung 3.17) untersucht. Wie schon bei *RPS2P32* wurde in allen Proben eine bialleliche Expression festgestellt. In den Proben aus Blut wurde zusätzlich noch eine SNaPshot-Analyse mit anschließendem Genescan (siehe Abschnitt 2.4.3.9, Rohdaten siehe Anhang Abschnitt A.9 Tabelle A.26) durchgeführt, wodurch die bialleliche Expression bestätigt wurde (Rate= 1,05 und Rate= 0,82, siehe auch Abbildung 3.16). Somit kann auch eine Regulation des differentiellen R-CGIs 9261_1_hg19 von *IGF2BP3* in Plazenta und Blut ausgeschlossen

werden.

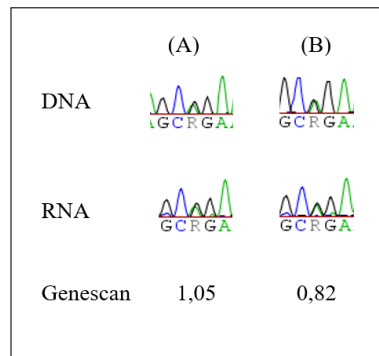


Abbildung 3.16: Expressionsanalyse von *IGF2BP3* in Blut

Diese Abbildung zeigt die Ergebnisse der Expressionsanalyse von *IGF2BP3* in Blut für zwei Individuen (A) und (B). Für jede Probe ist jeweils die DNA-(oben) und die RNA-(unten) Sequenz angegeben. In allen untersuchten Proben konnte eine bialleliche Expression festgestellt werden. Zusätzlich sind noch die Ergebnisse des Genescans (siehe Abschnitt 2.4.3.9) angegeben, die eine bialleliche Expression bestätigen.

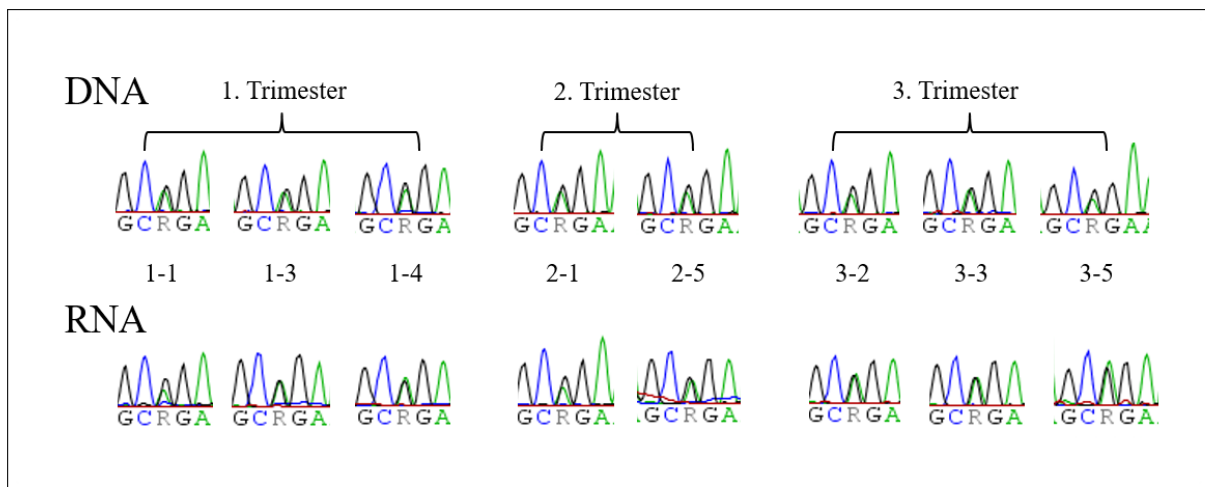


Abbildung 3.17: Expressionsanalyse von *IGF2BP3* in Plazentaprobe aus unterschiedlichen Trimestern

Die Abbildung zeigt die Ergebnisse der Expressionsanalyse von *IGF2BP3* in Plazentaprobe aus unterschiedlichen Trimestern. Für jede Probe ist jeweils die DNA-(oben) und die RNA-(unten) Sequenz angegeben. In allen untersuchten Proben konnte eine bialleliche Expression festgestellt werden. Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

4 Diskussion

4.1 Ist der *RB1/Rb1*-Locus einzigartig in seinen Eigenschaften?

Die meisten CGIs liegen in Wirbeltiergenomen am 5'-Ende (Promotorregion) oder im ersten Exon eines Gens und beinhalten sowohl Bindestellen für Transkriptionsfaktoren als auch für die RNA-Polymerase. In der Regel sind CGI unmethyliert. Über intronische CGIs ist allgemein wenig bekannt. Sie könnten die Expression des Wirtsgens modifizieren, eine alternative Transkriptionsstartseite beinhalten, zu einem Gen innerhalb des Introns des Wirtsgens gehören oder aber auch keine Funktion haben. Auch über den evolutionären Ursprung von intronischen CGIs ist nur wenig bekannt. Allerdings existieren mehrere intronische CGIs, wie z.B. das humane CpG85 (16634_1_hg19) des *RB1*-Gens, die durch Retrotransposition entstanden sind. Das CpG85 erwarb während der Evolution eine differentielle Methylierung, was die Hauptursache für die Prägung des *RB1*-Gens ist. Im Gegensatz hierzu beinhaltet das murine *Rb1*-Gen dieses CGI (CpG85) nicht und ist daher auch nicht geprägt (Kamber u. a., 2009a, 2013; Steenpass u. a., 2013).

In der vorliegenden genomweiten Studie wurde untersucht, ob der *RB1/Rb1*-Locus einzigartig bezüglich seiner oben genannten Eigenschaften ist. Gibt es weitere CGIs, die im Intron eines humanen Gens liegen, die nicht in der Maus vorkommen und eine differentielle Methylierung vorweisen? Desweiteren wurde der Zeitpunkt der Entstehung während der Evolution dieser intronischen CGIs, die nicht in der Maus vorhanden sind, analysiert. In der gesamten Studie dient die Identifikation des CpG85 des *RB1*-Gens (16634_1_hg19) als positive Kontrolle.

Durch die Berechnung der Lage von murinen und humanen CGIs in Abhängigkeit von einem übereinstimmenden Datensatz an proteinkodierenden Regionen (CCDS) für beide Organismen, wurde im humanen Genom eine deutlich höhere Prozentzahl an CGIs (mehr als doppelt so viele) als im murinen Genom festgestellt (siehe Tabelle 3.1). Der Mensch hat auch einen wesentlich höheren Anteil an intronischen CGIs als die Maus. Mittels Vergleich

der Sequenzen der intronischen CGIs von Mensch und Maus, wurde herausgefunden, dass 2.033 humane intronische CGIs nicht in der Maus und 470 murine intronische CGIs nicht im Menschen vorhanden sind. Dieses Ergebnis zeigt, dass in beiden evolutionären Abstammungslinien neue CGIs entstanden sind. Es ist durchaus möglich, dass mehr CGIs dieser Art sowohl im murinen als auch im humanen Genom existieren, da diese Studie nur auf Genen basiert, die in beiden Organismen und in dem CCDS Datensatz vorhanden sind.

Von diesen 2.033 human/nicht-murin intronischen CGIs, die in der vorliegenden Studie analysiert wurden, haben 104 CGIs eine hohe Sequenzähnlichkeit zu anderen Regionen im menschlichen Genom und mindestens 50% davon sind Teil einer Retrokopie. Die Mehrheit der 104 CGIs ist auch in den anderen untersuchten Mitgliedern der Superfamilie Hominoidea vorhanden (siehe Tabelle 3.5). Der Anteil, der auch in anderen Primaten vorkommenden CGIs, korreliert grob mit der evolutionären Verwandtschaft dieser Arten. Interessanterweise sind 13 CGIs im nahverwandten Genom des Schimpansen nicht vorhanden. Hierzu gehört auch das CGI 19100_1_hg19 (*PDXDC1*) das innerhalb der Studie ausführlich analysiert wurde (siehe unten). Im Buschbaby-Genom konnten nur sieben der 104 CGIs nachgewiesen werden. Dieses Ergebnis weist darauf hin, dass der Großteil der 104 CGIs nach der Spaltung der Unterordnungen Haplorrhini und Strepsirrhini entstanden sind, und unterstützt die Idee, dass CGIs Teil der *retrotranspositional explosion* sind. Laut Ohshima u. a. (2003) fand diese Explosion von prozessierten Pseudogenen und den repetitiven Alu-Elementen gleichzeitig in Haplorrhini 40-50 Millionen Jahre vor der Spaltung in die Unterordnungen Catarrhini und Platyrrhini statt. Als mögliche Mechanismen nennen Ohshima u. a. (2003) entweder eine unbekannte Änderung im zellulären Umfeld oder aber eine Weiterentwicklung der Zellteilung selbst, die die Zellteilung von Retrotransposons unterstützt.

Die Mehrheit der CGIs im humanen Genom ist unmethyliert. Jedoch sind ca. 75% der human/nicht-murin intronischen CGIs vollständig methyliert. Diese Ergebnisse weisen darauf hin, dass die CGIs durch einen Abwehrmechanismus der DNA vor Fremd-DNA methyliert wurden und dadurch möglicherweise funktionslos sind. Ein weiterer Abwehrmechanismus der DNA gegenüber Fremd-DNA ist z.B. die RNA-Interferenz (RNAi). Sowohl die Histonmethylierung als auch die RNAi können Retrokopien bzw. Retrogene kontrollieren und zu ihrer Abschaltung bzw. Stilllegung führen (Cam u. a., 2008). Nur etwa 15% der 104 human/nicht-murin intronischen CGIs sind unmethyliert. Das lässt darauf schließen, dass diese CGIs ihre ursprüngliche Funktion beibehalten oder aber in ihrer neuen Umgebung eine neue Funktion erworben haben. Fünf CGIs, einschließlich CpG85 des geprägten *RB1*-Gens, weisen ein mittleres Methylierungsniveau in humanen Monozyten

und auch anderen Gewebetypen auf (siehe Abschnitt 3.1.2, Abbildung 3.1, Tabelle 3.3 und Tabelle 3.4). Von diesen fünf CGIs scheinen vier durch Insertion einer Retrokopie entstanden zu sein, nur das CGI 20632_1_hg19 nicht (siehe Abbildung 3.5). Dieses CGI liegt im Intron 1 des *MYO1D*-Gens und ähnelt sehr einem CGI auf dem X-Chromosom. Es ist möglich, dass in diesem Bereich auf dem X-Chromosom ein Gen liegt oder aber eine Retrokopie, die jedoch noch nicht annotiert sind. Alle fünf CGIs sind nicht vollständig methyliert, daher könnte es möglich sein, dass diese CGIs eine neue Funktion erworben haben.

Wie schon in anderen Studien für den *RB1*-Locus beschrieben wurde, zeigt das CpG85 (16634_1_hg19) eine geprägte DNA-Methylierung (Kanber u. a., 2009b). Um die vier übrigen CGIs auf eine allelische Methylierung hin zu untersuchen, wurde die hochquantitative Hochdurchsatz-Bisulfitsequenzierung auf dem Roche/454 GS Junior verwendet. Es wurde DNA aus Monozyten von Individuen, die für einen ausgewählten SNP heterozygot sind, untersucht. Obwohl es nicht möglich war, das CGI 20632_1_hg19 (*MYO1D*) zu analysieren, konnte die Hypothese von einer allelischen Methylierung bei 15224_1_hg19 (*PARP11*) verworfen werden und teilweise allelische Methylierungsunterschiede bei CGI 14414_1_hg19 (*ASRGL1*) und CGI 19100_1_hg19 (*PDXDC1*) gezeigt werden. Die Methylierungsunterschiede der einzelnen Allele beim *PDXDC1*-Locus waren deutlich stärker als die beim *ASRGL1*-Locus, aber nicht so stark wie beim *RB1*-Locus. Zusätzlich weisen die Daten darauf hin, dass die beobachteten allelischen Methylierungsunterschiede an den beiden Loci (*ASRGL1* und *PDXDC1*) nicht elternspezifisch sondern eher sequenzspezifisch sind (siehe Abbildungen 3.2 und 3.3 und Tabellen A.11, A.12, A.13 und A.14).

Darüber hinaus wurden die Ergebnisse dieser Studie mit denen von Court u. a. (2014) verglichen. Court und Kollegen führten eine genomweite Suche nach geprägten Genen durch und beschrieben 21 neue differentiell-methylierte Regionen (DMRs). Von diesen 21 DMRs sind 15 Placenta-spezifisch und wurden daher nicht in der vorliegenden Studie, die auf Monozyten basiert, gefunden. Die anderen sechs neuen DMRs, die eine mittlere Methylierung in fünf unterschiedlichen Gewebetypen (Blut, Gehirn, Leber, Muskel und Niere) aufwiesen, wurden ebenfalls durch diese Analysen nicht gefunden. Vier DMRs (*PPIEL*, *WDR27*, *HTR5A* und *CXORF56*) sind nur CpG-reiche Regionen und im UCSC Browser nicht als CpG-Inseln annotiert. Da die vorliegende Studie auf Daten des UCSC Browsers basiert, konnten die vier DMRs nicht entdeckt werden. Die zwei übrigen DMRs, die Court u. a. (2014) noch gefunden haben, liegen zwar in einem Intron, wurden jedoch in den vorliegenden Analysen nicht gefunden. *NHP2L1* weist keinen weiteren Treffer im humanen Genom auf und *WRB* liegt in unserem Datensatz nicht in einem Intron. In Court

u. a. (2014) wurde auch der geprägte *RB1*-Locus gefunden, jedoch kein weiteres CGI, was in der vorliegenden Studie untersucht wurde. Ebenso beobachteten Court u. a. (2014) differentielle Methylierung des CGIs 19100_1_hg19 (*PDXDC1*). Basierend auf Daten von uniparentalen Disomien, wurde dieser Locus jedoch als geprägt ausgeschlossen. Die Methylierung an diesem Locus scheint abhängig von der Sequenz zu sein. Dementsprechend ergänzen sich diese beiden Studien mit unterschiedlichen Zielen.

Innerhalb der vorliegenden Studie wurde kein weiterer Locus mit denselben Eigenschaften wie der *RB1/Rb1*-Locus gefunden, daher ist der *RB1/Rb1*-Locus einzigartig in seinen Eigenschaften.

4.2 Analyse von Retrokopie-assoziierten CGIs

Dieser Teil der Arbeit beschäftigt sich mit dem Methylierungsgrad aller CGIs, die mit einer Retrokopie assoziiert sind, der Methylierung der genomischen Umgebung der Retrokopie, sowie mit der Methylierung des ancestralen Gens (siehe Abbildung 3.6). Grundlage dieser Untersuchungen sind genomweite und methylierungsspezifische Sequenzdaten von humanen Monozyten. Wie auch schon in 4.1 erläutert, liegen die meisten CGIs in der Promotorregion bzw. im ersten Exon eines Gens und sind unmethyliert. Offenbar schützt die Bindung von Transkriptionsfaktoren diese Bereiche vor Methylierung (Stadler u. a., 2011). Fehlende Methylierung bei CGIs hat im Laufe der Evolution zu einer Erhaltung von Cytosinresten geführt. Dies steht im Gegensatz zu anderen Regionen im menschlichen Genom, in denen die Desaminierung von Methylcytosin zu C→T-Transitionen geführt hat. Wie in Abschnitt 1.5 beschrieben, bestehen Retrokopien hauptsächlich aus den Exons eines Gens und nicht aus Introns oder den Promotorregionen eines Gens. Jedoch ist es zusätzlich möglich, dass im Laufe der Evolution neue CGIs durch Retrotransposition entstanden sind. All diese Fakten erklären zusammen, warum nur 10% aller Retrokopien (1.319 von 13.173) mit einem CGI assoziiert sind.

Wie bereits oben beschrieben schützt die Bindung von Transkriptionsfaktoren an CGIs vor Methylierung (Stadler u. a., 2011). Dies ist für die meisten menschlichen Gene der Fall. Im Gegensatz hierzu haben Retrokopien oft keine Funktion oder aber sind nur schwach exprimiert (siehe Abschnitt 1.5). Dies erklärt, dass nur etwa ein Drittel aller R-CGIs unmethyliert ist. Desweiteren ist es möglich, dass Sequenzen, die durch Retrotransposition

entstanden sind, dem Abwehrmechanismus des menschlichen Genoms gegenüber Fremd-DNA unterliegen und daher methyliert werden. Unmethylierte CGIs könnten diesen Mechanismus verloren oder aber einen Mechanismus gegen diese Methylierung durch die Retrotransposition hinzugewonnen haben.

Bei der Methylierung von R-CGIs eines anzestralen Gens konnten keine Regelmäßigkeiten bzw. Gemeinsamkeiten beobachtet werden, sondern es wurden große Unterschiede in der Methylierung festgestellt (siehe Abbildung 3.7). Dieses Ergebnis zeigt, dass der Methylierungsstatus nicht nur, bzw. nicht ausschließlich, durch die retrotransponierte Sequenz bestimmt wird. Auch konnte kein Zusammenhang zwischen der Methylierung eines R-CGIs und der Expression der Retrokopie festgestellt werden, bzw. der genomischen Lage eines R-CGIs.

Es konnte jedoch ein starker Zusammenhang zwischen dem Methylierungsgrad von R-CGIs und der genomischen Umgebung festgestellt werden. Das Vorkommen von methylierten CGIs in ausschließlich methylierter Umgebung (siehe Cluster 2 in Abbildung 3.8) weist auf einen starken Effekt der Methylierung der Integrationsstelle auf den Methylierungsgrad des R-CGIs hin (Günzburg u. Groner, 1984). Ein Einfluss von möglichen repetitiven Sequenzen in der unmittelbaren Umgebung wurde durch Analysen ausgeschlossen. Im Hinblick auf geprägte Gene gibt es einige Beispiele, in dem die Retrokopie den Methylierungsstatus durch seine Integrationsstelle erworben hat. Beispiele hierfür sind *NDN*, entstanden durch eine Retrotransposition eines Gens vom X-Chromosom, und *NPAP1*, entstanden durch eine Retrotransposition eines Gens von Chromosom 7 (Chai u. a., 2001; Rapkins u. a., 2006; Neumann u. a., 2014). Beide Gene gelangten durch Retrotransposition in die Region 15q11q13, welche als geprägt bekannt ist, und wurden an dieser Stelle im Genom anschließend von dem *cis*-agierenden *Imprintingcenter* kontrolliert. *NDN* ist auf der Liste der 1.319 R-CGIs, fällt jedoch bei der anschließenden Methylierungsanalyse heraus, da mehr als 40% der Reads eine gemischte Methylierung aufweisen. Diese gemischte Methylierung ist ungewöhnlich für ein geprägtes Gen, wurde aber bei *NDN* schon mehrfach beobachtet (Court u. a., 2014). Dies kann unter anderem mit der Größe des CGIs (1 kb) zusammenhängen, welches eventuell größer ist als die eigentliche DMR. *NPAP1* und andere geprägte Retrokopien wie *PEG10* sind nicht auf der Liste der 1.319 RO-CGIs, da diese innerhalb des UCSC Browsers nicht als Retrokopien klassifiziert sind (Suzuki u. a., 2007). Der Ursprung von *NPAP1* wurde erst vor kurzem von Neumann u. a. (2014) beschrieben.

Im Gegensatz zur eben beschriebenen Theorie ist es jedoch auch möglich, dass sich die Methylierung der Retrokopie durch Abwehrmechanismen in die umliegende Region

ausgebreitet hat (Jähner u. Jaenisch, 1985). Sollte dies der Fall sein, so lässt sich der starke Zusammenhang zwischen Methylierung von R-CGIs und der Methylierung der genomischen Region als *de novo* Methylierung infolge der Retrotransposition erklären. Es lässt sich an dieser Stelle jedoch keine Aussage darüber machen, ob das Vorkommen von methylierten CGIs in methylierten Regionen und das Vorkommen von unmethylierten CGIs in unmethylierten Regionen zeigt, dass Retrokopien mit gleicher Wahrscheinlichkeit in unmethylierte und methylierte Regionen integriert werden.

Interessanterweise wurden R-CGIs mit einem mittleren Methylierungsgrad (zwischen 20% und 80%) ausschließlich in Regionen gefunden, die mehr als 40% methyliert sind. Bei reinen Zellpopulationen, wie den Monozyten, kann ein mittleres Methylierungslevel verschiedene Ursachen haben, wie z.B. zelluläre Heterogenität, uneinheitliche Methylierung beider Allele, ein Anstieg bzw. Gefälle der Methylierung innerhalb eines CGIs, allelspezifische Methylierung infolge einer in *cis*-agierenden Variante, und geprägte Methylierung. Unter den 195 R-CGIs mit einem mittleren Methylierungslevel konnten nur zwei CGIs mit einer geprägten Methylierung identifiziert werden: Das bereits publizierte CpG85 des *RB1*-Gens (siehe Abschnitt 1.6), das mit einer Retrokopie des *PPP1R26*-Gens assoziiert ist und das in dieser Arbeit beschriebene R-CGI 9261_1_hg19, das mit dem *RPS2P32*-Gen assoziiert ist. In beiden Fällen ist die Retrokopie in eine nicht-geprägte Region integriert worden und hat eine Methylierung auf dem maternalen Allel erworben. Die genomische Umgebung von *PPP1R26P1* besitzt einen Methylierungsgrad von 79%. Das CpG85 selbst besitzt ein alternatives Startexon, das auf das dritte Exon des *RB1*-Gens gespleißt wird. Dieses alternative Transkript des *RB1*-Gens wird nur vom paternalen Allel transkribiert. Die Integration dieser Retrokopie ist verantwortlich für die genomische Prägung des *RB1*-Gens und der damit verbundenen präferentiell maternalen Expression (Kanber u. a., 2009a, 2013).

Das *RPS2P32*-Gen auf Chromosom 7 ist durch eine Retrokopie des *RPS2*-Gens auf Chromosom 16 entstanden (siehe Abbildung 3.9). Die genomische Umgebung von *RPS2P32* weist eine Methylierung von 90% auf. Das ancestrale *RPS2*-Gen enthält ebenfalls ein CGI, das jedoch unmethyliert ist. Die Ergebnisse der Analyse von unterschiedlichen Histonmodifikationen (siehe Abschnitt 3.2.3.1 und Abbildung 3.10) zeigt, wie auch bei anderen geprägten Genen, neben aktiven Promotoren und Enhancern auch Heterochromatin. Eine gleichzeitige Aktivität von aktiven und reprimierten Modifikationen kann auch die unterschiedlichen Allele widerspiegeln. Dies könnte ein weiterer Hinweis auf eine mögliche Prägung des *RPS2P32*-Gen sein.

In diesem Teil der Arbeit wurde durch Oozyten-spezifische Methylierung und Methylierung

des maternalen Allels in Blut gezeigt, dass das R-CGI 9261_1_hg19 eine allelspezifische Methylierung aufweist. Desweiteren zeigen vier von sechs untersuchten Patienten mit einem MLID einen Verlust der Methylierung des CGIs 9261_1_hg19. Alle eben genannten Faktoren weisen darauf hin, dass das CGI eine geprägte Methylierung trägt. Jedoch konnte in Blut (siehe Abbildung 3.14) nur eine biallelische Expression von *RPS2P32* festgestellt werden, wobei auf dem Gegenstrang eine leichte Verschiebung der Expression zugunsten des paternalen (unmethylierten) Allels festgestellt werden konnte. Dieses Ergebnis ist unerwartet, da im UCSC Browser kein *antisense*-Transkript angezeigt wird. Hannula-Jouppi u. a. (2014) konnten in Patienten mit einer maternalen uniparentalen Disomie 7 (matUPD7) eine Hypomethylierung des *RPS2P32*-Gens feststellen. Desweiteren fanden sie auch keine signifikanten Unterschiede in der Expression zwischen Patienten mit einer matUPD7 und Kontrollen. Dieses Ergebnis schließt *RPS2P32* auf den ersten Blick als ein neues geprägtes Gen aus, jedoch ist es durchaus möglich, dass *RPS2P32* in einem anderen Gewebe monoallelisch exprimiert wird.

Da es nicht möglich war, eine monoallelische Expression für *RPS2P32* nachzuweisen, wurden die benachbarten Gene von *RPS2P32*, *TRA2A* und *IGF2BP3*, untersucht (siehe Abbildung 3.9). *IGF2BP3* kodiert für ein Protein, das die mRNA-Stabilität und mRNA-Translation von *IGF2* reguliert. *IGF2* ist ein geprägtes Gen, das in der menschlichen Entwicklung und dem Wachstum involviert ist. Expression des paternalen Allels des *IGF2*-Gens fördert das Wachstum (Liao u. a., 2005, 2011; Suvasini u. a., 2011). *IGF2BP3* könnte als Wachstumsförderer ein sehr guter Kandidat für ein neues geprägtes Gen sein. *IGF2BP3* besitzt zwei CGIs (siehe Abbildung 3.9) und im ersten Schritt wurde die Methylierung dieser CGIs untersucht. Beide CGIs sind jedoch unmethyliert (siehe Abbildung 3.15). Im nächsten Schritt wurde die Expression von *IGF2BP3* untersucht. *IGF2BP3* wird in Blut biallelisch exprimiert, was auch schon durch Monk u. a. (2002) gezeigt wurde. Li u. a. (2014) zeigten, dass *IGF2BP3* in Plazenta stark exprimiert wird, wobei die Expression im ersten Trimester einer Schwangerschaft am stärksten ist und zum dritten Trimester hin abnimmt. Desweiteren stellten sie fest, dass die Expression von *IGF2BP3* in Plazenten von Patientinnen mit einer Präeklampsie stark verringert ist, einhergehend mit einem verzögerten intrauterinen Wachstum. Daher wurde ebenfalls die Expression von *IGF2BP3* in Plazenten unterschiedlicher Schwangerschaftstrimester untersucht. Die Untersuchungen ergaben in allen Proben eine biallelische Expression von *IGF2BP3* (siehe Abbildung 3.17). Dennoch ist es wie auch für *RPS2P32* möglich, dass eine monoallelische Expression von *IGF2BP3* beschränkt ist auf einen bestimmten Gewebe- oder Zelltyp, der nicht untersucht wurde. Ein gutes Beispiel für diese Annahme

ist das *UBE3A*-Gen, welches in fast allen Geweben biallelisch exprimiert wird außer im Gehirn, in dem es monoallelisch vom maternalen Allel exprimiert wird (Rougeulle u. a., 1997). Die Transkription von *UBE3A* wird durch ein *Antisense*-Transkript kontrolliert, das seinerseits von einer *Imprinting*kontrollregion (ICR) reguliert wird (Chamberlain u. Brannan, 2001; Landers u. a., 2005; Royo u. Cavaillé, 2008; Meng u. a., 2012). Diese ICR ist in allen Geweben paternal unmethyliert und maternal methyliert (Horsthemke u. Wagstaff, 2008).

5 Zusammenfassung

CpG-Inseln sind, im Vergleich zum übrigen Genom, DNA-Regionen mit einem hohen GC-Gehalt und einer hohen Frequenz an CpG-Dinukleotiden. Sie befinden sich meistens am 5'-Ende von Genen und sind in der Regel unmethyliert. Ausnahme sind durch X-Inaktivierung und genomische Prägung inaktivierte Gene. Die genomische Prägung des humanen *RB1*-Gens beruht z.B. auf einem differentiell methylierten CGI im Intron 2 des *RB1*-Gens. Dieses CGI ist Teil einer durch Retrotransposition eingefügten Sequenz des *PPP1R26*-Gens von Chromosom 9 und hat sich zu einem differentiell methylierten alternativen *RB1*-Promotor entwickelt. Ziel dieser Arbeit ist es weitere geprägte Loci im Humangenom mit denselben Eigenschaften, wie denen des *RB1*-Locus, zu finden, wenn möglich den evolutionären Zeitpunkt der Entstehung dieser CGIs zu bestimmen und darüber hinaus alle CGIs, die mit einer Retrokopie überlappen auf genomische Prägung hin zu untersuchen.

Zusammenfassend wurde im ersten Teil dieser Arbeit herausgefunden, dass das humane Genom nicht nur mehr CGIs als das murine Genom besitzt, sondern auch das Verhältnis von ausschließlich intronischen CGIs mehr als doppelt so hoch ist (7,7% humane intronische CGIs zu 3,5% murinen intronischen CGIs). Mindestens 2.033 humane intronische CGIs sind nicht im murinen Genom vorhanden. Von diesen haben 104 CGIs Sequenzähnlichkeiten zu anderen Regionen im humanen Genom und mindestens 45 sind assoziiert mit einer Retrokopie.

Insgesamt konnte von 86 der 104 human/nicht-murin intronischen CGIs der evolutionäre Ursprung näher bestimmt werden. Im Galago-Genom (Buschbaby) konnten nur sieben dieser CGIs nachgewiesen werden. Dies weist darauf hin, dass der Großteil der 104 CGIs nach der Spaltung der Unterordnungen Haplorrhini und Strepsirrhini entstanden sind, und unterstützt die Idee, dass CGIs Teil der *retrotranspositional explosion* sind.

Die meisten der human/nicht-murin intronischen CGIs sind methyliert (~75%) oder unmethyliert (~15%). Nur 14 CGIs zeigten ein mittleres Methylierungsniveau zwischen 20% und 80% in Monozyten und auch anderen Geweben, wovon nur fünf (14414_1_hg19, 15224_1_hg19, 16634_1_hg19, 19100_1_hg19 und 20632_1_hg19), einschließlich des

intronischen CpG85 des *RB1*-Gens (16634_1_hg19), weitere Kriterien für eine mögliche differentielle Methylierung erfüllten. Diese CGIs sind, bis auf CGI 20632_1_hg19, alle durch Retrotransposition entstanden. Mit Hilfe der hochquantitativen Hochdurchsatz-Bisulfitsequenzierung konnten bei CGI 14414_1_hg19 (*ASRGL1*) und 19100_1_hg19 (*PDXDC1*) allelische Methylierungsunterschiede gezeigt werden. Im Gegensatz zu der geprägten Methylierung des CpG85 des *RB1*-Gens, waren die allelischen Methylierungsunterschiede bei *ASRGL1* und *PDXDC1* jedoch deutlich geringer, so dass ihre differentielle Methylierung vermutlich sequenzspezifisch ist. Die Hypothese von einer allelspezifischen Methylierung von CGI 15224_1_hg19 (*PARP11*) konnte verworfen werden.

Die Ergebnisse der vorliegenden Arbeit zeigen, dass der *RB1/Rb1*-Locus einzigartig in seinen Eigenschaften ist. Es gibt jedoch weitere CGIs, die allelische Methylierungsunterschiede aufweisen, die aber nicht elternspezifisch sind, sondern abhängig von der DNA-Sequenz.

Im zweiten Teil der vorliegenden Arbeit konnte ermittelt werden, dass insgesamt nur 10% (1.319 von 13.173) aller Retrokopien mit einem CGI assoziiert sind (R-CGIs). Der Großteil der R-CGIs war entweder vollständig methyliert (588) oder vollständig unmethyliert (384). Bei 152 R-CGIs konnte die Methylierung aufgrund schlechter genomischer Abdeckung nicht ermittelt werden. Ein mittleres Methylierungsniveau (zwischen 20% und 80%) wurde bei 195 dieser sogenannten R-CGIs festgestellt, wobei neben der geprägten Methylierung des CpG85 des *RB1*-Gens auch eine geprägte Methylierung des R-CGIs 9261_1_hg19 des *RPS2P32*-Gens identifiziert werden konnte. Methylierung in Oozyten, keine Methylierung in Sperma, Methylierung nur des maternalen Allels in DNA aus Blut und ein Verlust der Methylierung in vier von sechs Patienten mit einem MLID weisen stark auf eine genomische Prägung hin. Untersuchungen der Expression in Blut zeigten jedoch eine biallelische Expression des *RPS2P32*-Gens. Daher ist es möglich, dass *RPS2P32* in einem anderen Gewebe monoallelisch exprimiert wird. Auch konnte keine monoallelische Expression des benachbarten *IGF2BP3*-Gens, weder in Blut noch in Plazenta, nachgewiesen werden. Auch in diesem Fall kann eine monoallelische Expression von *IGF2BP3* in anderen Geweben nicht ausgeschlossen werden.

Es konnte ein starker Zusammenhang zwischen dem Methylierungsgrad eines R-CGIs und der genomischen Umgebung festgestellt werden. Einerseits ist es möglich, dass die Methylierung der Integrationsstelle einen starken Effekt auf die Methylierung des R-CGIs hat, wobei ein potentieller Einfluss von repetitiven Sequenzen ausgeschlossen werden konnte. Andererseits kann sich auch die Methylierung der Retrokopie z.B. durch Abwehrmechanismen ausgebreitet haben.

Literaturverzeichnis

- [Bannister u. Kouzarides 2011] BANNISTER, AJ. ; KOUZARIDES, T.: Regulation of chromatin by histone modifications. In: *Cell research* 21 (2011), Nr. 3, S. 381–395
- [Barlow u. Bartolomei 2014] BARLOW, DP. ; BARTOLOMEI, MS.: Genomic imprinting in mammals. In: *Cold Spring Harbor perspectives in biology* 6 (2014), Nr. 2, S. a018382. <http://dx.doi.org/10.1101/cshperspect.a018382>. – DOI 10.1101/cshperspect.a018382
- [Barlow u. a. 1991] BARLOW, DP. ; STÖGER, R. ; HERRMANN, BG. ; SAITO, K. ; SCHWEIFER, N.: The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. In: *Nature* 349 (1991), Nr. 6304, S. 84–87
- [Bartolomei u. Ferguson-Smith 2011] BARTOLOMEI, MS. ; FERGUSON-SMITH, AC.: Mammalian genomic imprinting. In: *Cold Spring Harbor perspectives in biology* 3 (2011), Nr. 7, S. a002592
- [Beygo u. a. 2013] BEYGO, J. ; CITRO, V. ; SPARAGO, A. ; CRESCENZO, A. D. ; CERRATO, F. ; HEITMANN, M. ; RADEMACHER, K. ; GUALA, A. ; ENKLAAR, T. ; ANICHINI, C. ; SILENGO, M. C. ; GRAF, N. ; PRAWITT, D. ; CUBELLIS, MV. ; HORSTHEMKE, B. ; BUITING, K. ; RICCIO, A.: The molecular function and clinical phenotype of partial deletions of the IGF2/H19 imprinting control region depends on the spatial arrangement of the remaining CTCF-binding sites. In: *Hum Mol Genet* 22 (2013), Nr. 3, S. 544–557. <http://dx.doi.org/10.1093/hmg/dds465>. – DOI 10.1093/hmg/dds465
- [Bird 2002] BIRD, A.: DNA methylation patterns and epigenetic memory. In: *Genes Dev* (2002), Nr. 16, S. 6–21. <http://dx.doi.org/10.1101/gad.947102>. – DOI 10.1101/gad.947102
- [Buiting u. a. 2001] BUITING, K. ; BARNICOAT, A. ; LICH, C. ; PEMBREY, M. ; MALCOLM, S. ; HORSTHEMKE, B.: Disruption of the Bipartite Imprinting Center in a Family with Angelman Syndrome. In: *Am J Hum Genet* 68 (2001), Nr. 5, S. 1290–1294. <http://dx.doi.org/10.1086/320120>. – DOI 10.1086/320120
- [Cam u. a. 2008] CAM, HP. ; NOMA, K. ; EBINA, H. ; LEVIN, HL. ; GREWAL, SI.: Host genome surveillance for retrotransposons by transposon-derived proteins. In: *Nature* 451 (2008), Nr. 7177, S. 431–436
- [Cedar u. Bergman 2009] CEDAR, H. ; BERGMAN, Y.: Linking DNA methylation and histone modification: patterns and paradigms. In: *Nature Reviews Genetics* 10 (2009), Nr. 5, S. 295–304

- [Chai u. a. 2001] CHAI, J-H. ; LOCKE, DP. ; OHTA, T. ; GREALLY, JM. ; NICHOLLS, RD.: Retrotransposed genes such as *Frat3* in the mouse Chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site. In: *Mammalian Genome* 12 (2001), Nr. 11, S. 813–821
- [Chamberlain u. Brannan 2001] CHAMBERLAIN, SJ. ; BRANNAN, CI.: The Prader–Willi syndrome imprinting center activates the paternally expressed murine *Ube3a* antisense transcript but represses paternal *Ube3a*. In: *Genomics* 73 (2001), Nr. 3, S. 316–322
- [Court u. a. 2014] COURT, F. ; TAYAMA, C. ; ROMANELLI, V. ; MARTIN-TRUJILLO, A. ; IGLESIAS-PLATAS, I. ; OKAMURA, K. ; SUGAHARA, N. ; SIMON, C. ; MOORE, H. ; HARNESS, JV. ; KEIRSTEAD, H. ; SANCHEZ-MUT, JV. ; KANEKI, E. ; LAPUNZINA, P. ; SOEJIMA, H. ; WAKE, N. ; ESTELLER, M. ; OGATA, T. ; HATA, K. ; NAKABAYASHI, K. ; MONK, D.: Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation independent mechanism of establishment. In: *Genome Res* 4 (2014), Nr. 24, S. 554–569. <http://dx.doi.org/10.1101/gr.164913.113>. – DOI 10.1101/gr.164913.113
- [Cowley u. Oakey 2010] COWLEY, M. ; OAKEY, RJ.: Retrotransposition and genomic imprinting. In: *Briefings in functional genomics* 9 (2010), Nr. 4. <http://dx.doi.org/10.1093/bfpg/elq015>. – DOI 10.1093/bfpg/elq015
- [Dean 2014] DEAN, W.: DNA methylation and demethylation: A pathway to gametogenesis and development. In: *Mol Reprod Dev* (2014), Nr. 81, S. 113–125. <http://dx.doi.org/10.1002/mrd.22280>. – DOI 10.1002/mrd.22280
- [Edwards u. Ferguson-Smith 2007] EDWARDS, CA. ; FERGUSON-SMITH, AJ.: Mechanisms regulating imprinted genes in clusters. In: *Current Opinion in Cell Biology* (2007), Nr. 19, S. 281–289. <http://dx.doi.org/10.1016/j.ceb.2007.04.013>. – DOI 10.1016/j.ceb.2007.04.013
- [Esnault u. a. 2000] ESNAULT, C. ; MAESTRE, J. ; HEIDMANN, T.: Human LINE retrotransposons generate processed pseudogenes. In: *Nat Genet* (2000), Nr. 24, S. 363–367. <http://dx.doi.org/10.1038/74184>. – DOI 10.1038/74184
- [Fatemi u. a. 2005] FATEMI, M. ; PAO, MM. ; JEONG, S. ; GAL-YAM, E. N. ; EGGER, G. ; WEISENBERGER, DJ. ; JONES, PA.: Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. In: *Nuc Acids Res* 20 (2005), Nr. 33. <http://dx.doi.org/10.1093/nar/gni180>. – DOI 10.1093/nar/gni180
- [Ferguson-Smith 2011] FERGUSON-SMITH, AC.: Genomic imprinting: the emergence of an epigenetic paradigm. In: *Nature Reviews Genetics* 12 (2011), Nr. 8, S. 565–575
- [Friend u. a. 1986] FRIEND, SH. ; BERNARDS, R. ; ROGELJ, S. ; WEINBERG, RA. ; RAPAPORT, JM. ; ALBERT, DM. ; DRYJA, TP.: A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. In: *Nature* (1986), Nr. 323, S. 643–646. <http://dx.doi.org/10.1038/323643a0>. – DOI 10.1038/323643a0

- [Frommer u. a. 1992] FROMMER, M. ; McDONALD, LE. ; MILLAR, DS. ; COLLIS, CM. ; WATT, F. ; GRIGG, GW. ; MOLLOY, PL. ; PAUL, CL.: A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. In: *Proceedings of the National Academy of Sciences* 89 (1992), Nr. 5, S. 1827–1831
- [Fuks 2005] FUKS, F.: DNA methylation and histone modifications: teaming up to silence genes. In: *Current opinion in genetics & development* 15 (2005), Nr. 5, S. 490–495
- [Gardiner-Garden u. Frommer 1987] GARDINER-GARDEN, M. ; FROMMER, M.: CpG islands in vertebrate genomes. In: *J Mol Biol* 196 (1987), Nr. 2, S. 261–282
- [Greer u. Shi 2012] GREER, EL. ; SHI, Y.: Histone methylation: a dynamic mark in health, disease and inheritance. In: *Nature Reviews Genetics* 13 (2012), Nr. 5, S. 343–357
- [Grothaus u. a. 2016] GROTHAUS, K. ; KANBER, D. ; GELLHAUS, A. ; MIKAT, B. ; KOLAROVA, J. ; SIEBERT, R. ; WIECZOREK, D. ; HORSTHEMKE, B.: Genome-wide methylation analysis of retrocopy-associated CpG islands and their genomic environment. In: *Epigenetics* (2016). <http://dx.doi.org/10.1080/15592294.2016.1145330>. – DOI 10.1080/15592294.2016.1145330
- [Guenet 2005] GUENET, JL.: The mouse genome. In: *Genome Res* 12 (2005), Nr. 15, S. 1729–1740. <http://dx.doi.org/10.1101/gr.3728305>. – DOI 10.1101/gr.3728305
- [Günzburg u. Groner 1984] GÜNZBURG, WH. ; GRONER, B.: The chromosomal integration site determines the tissue-specific methylation of mouse mammary tumour virus proviral genes. In: *The EMBO journal* 3 (1984), Nr. 5, S. 1129
- [Haaf 2006] HAAF, T.: Methylation dynamics in the early mammalian embryo: implications of genome reprogramming defects for development. In: *DNA Methylation: Development, Genetic Disease and Cancer*. Springer, 2006, S. 13–22
- [Hackett u. Surani 2013] HACKETT, JA. ; SURANI, M. A.: DNA methylation dynamics during the mammalian life cycle. In: *Phil Trans R Soc B* 386 (2013). <http://dx.doi.org/10.1098/rstb.2011.0328>. – DOI 10.1098/rstb.2011.0328
- [Haig 1997] HAIG, D.: Parental antagonism, relatedness asymmetries, and genomic imprinting. In: *Proceedings of the Royal Society of London B: Biological Sciences* 264 (1997), Nr. 1388, S. 1657–1662
- [Hall 1990] HALL, JG.: Genomic imprinting: review and relevance to human diseases. In: *American journal of human genetics* 46 (1990), Nr. 5, S. 857
- [Hanna u. Kelsey 2014] HANNA, CW. ; KELSEY, G.: The specification of imprints in mammals. In: *Heredity* 113 (2014), Nr. 2, S. 176–183. <http://dx.doi.org/10.1038/hdy.2014.54>. – DOI 10.1038/hdy.2014.54
- [Hannula-Jouppi u. a. 2014] HANNULA-JOUPPI, K. ; MUURINEN, M. ; LIPSANEN-NYMAN, M. ; REINIUS, LE. ; EZER, S. ; GRECO, D. ; KERE, J.: Differentially methylated regions in maternal and paternal uniparental disomy for chromosome 7. In: *epigenetics* 9 (2014), Nr. 3, S. 351–365

- [Hartigan u. Wong 1979] HARTIGAN, JA. ; WONG, MA.: Algorithm AS 136: A k-means clustering algorithm. In: *Applied statistics* (1979), S. 100–108
- [Holliday u. Pugh 1996] HOLLIDAY, R. ; PUGH, JE.: DNA modification mechanisms and gene activity during development. In: *Cold Spring Harbor Monograph Archive* 32 (1996), S. 639–645
- [Horsthemke 2010] HORSTHEMKE, B.: Mechanisms of imprint dysregulation. In: *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* Bd. 154 Wiley Online Library, 2010, S. 321–328
- [Horsthemke 2014] HORSTHEMKE, B.: In Brief: Genomic imprinting and imprinting diseases. In: *J Pathol* 232 (2014), S. 458–487. <http://dx.doi.org/10.1002/path.4326>. – DOI 10.1002/path.4326
- [Horsthemke u. Buiting 2008] HORSTHEMKE, B. ; BUITING, K.: Genomic imprinting and imprinting defects in humans. In: *Advances in genetics* 61 (2008), S. 225–246
- [Horsthemke u. a. 2013] HORSTHEMKE, B. ; KANBER, D. ; BEYGO, J. ; WAGNER, N. ; RAHMANN, S. ; BUITING, K.: *DNA Methylation Analysis Using Deep Bisulfite Amplicon Sequencing on the GS Junior System*. 2013 (Roche Sequencing Application Note)
- [Horsthemke u. Wagstaff 2008] HORSTHEMKE, B. ; WAGSTAFF, J.: Mechanisms of imprinting of the Prader–Willi/Angelman region. In: *American journal of medical genetics Part A* 146 (2008), Nr. 16, S. 2041–2052
- [Ideraabdullah u. a. 2008] IDERAABDULLAH, FY. ; VIGNEAU, S. ; BARTOLOMEI, MS.: Genomic imprinting mechanisms in mammals. In: *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 647 (2008), Nr. 1, S. 77–85
- [Illingworth u. Bird 2009] ILLINGWORTH, RS. ; BIRD, AP.: CpG islands - a rough guide. In: *FEBS letter* 583 (2009), Nr. 11, S. 1713–1720. <http://dx.doi.org/10.1016/j.febslet.2009.04.012>. – DOI 10.1016/j.febslet.2009.04.012
- [Jähner u. Jaenisch 1985] JÄHNER, D. ; JAENISCH, R.: Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity. (1985)
- [Jenuwein u. Allis 2001] JENUWEIN, T. ; ALLIS, CD.: Translating the histone code. In: *Science* 293 (2001), Nr. 5532, S. 1074–1080
- [Jin u. a. 2014] JIN, C. ; LU, Y. ; JELINEK, J. ; LIANG, S. ; ESTECIO, MRH. ; BARTON, MC. ; ISSA, JPJ.: TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. In: *Nucleic acids research* 42 (2014), Nr. 11, S. 6956–6971
- [Jones 2012] JONES, PA.: Functions of DNA methylation: islands, start sites, gene bodies and beyond. In: *Nat Rev Genet* 3 (2012), Nr. 11, S. 484–492
- [Jurka 2000] JURKA, J.: Repbase update: a database and an electronic journal of repetitive elements. In: *Trends in genetics* 16 (2000), Nr. 9, S. 418–420

- [Kaessmann 2010] KAESSMANN, H.: Origins, evolution, phenotypic impact of new genes. In: *Genome Res* (2010), Nr. 20, S. 1313–1326. <http://dx.doi.org/10.1101/gr.101386.109>. – DOI 10.1101/gr.101386.109
- [Kaessmann u. a. 2009] KAESSMANN, H. ; VINCKENBOSCH, N. ; LONG, M.: RNA-based gene duplication: mechanistic and evolutionary insights. In: *Nat Rev Genet* (2009), Nr. 10, S. 19–31. <http://dx.doi.org/10.1038/nrg2487>. – DOI 10.1038/nrg2487
- [Kanber u. a. 2009a] KANBER, D. ; BERULAVA, T. ; AMMERPOHL, O. ; MITTER, D. ; RICHTER, J. ; SIEBERT, R. ; HORSTHEMKE, B. ; LOHMANN, D. ; BUITING, K.: The human retinoblastoma gene is imprinted. In: *PLoS One* 5 (2009), Nr. 12. <http://dx.doi.org/10.1371/journal.pgen.1000790>. – DOI 10.1371/journal.pgen.1000790
- [Kanber u. a. 2013] KANBER, D. ; BUITING, K. ; ROOS, C. ; GROMOLL, J. ; KAYA, S. ; HORSTHEMKE, B. ; LOHMANN, D.: The origin of the RB1 imprint. In: *PLoS One* 8 (2013), Nr. 11. <http://dx.doi.org/10.1371/journal.pone.0081502>. – DOI 10.1371/journal.pone.0081502
- [Kanber u. a. 2009b] KANBER, D. ; GILTAY, J. ; WIECZOREK, D. ; ZOGEL, C. ; HOCHSTENBACH, R. ; CALIEBE, A. ; KUECHLER, A. ; HORSTHEMKE, B. ; BUITING, K.: A paternal deletion of MKRN3, MAGEL2 and NDN does not result in Prader–Willi syndrome. In: *European Journal of Human Genetics* 17 (2009), Nr. 5, S. 582–590
- [Kent 2002] KENT, WJ.: BLAT-the BLAST-like alignment tool. In: *Genome Res* 12 (2002), Nr. 4, S. 656–664
- [Kivelä 2009] KIVELÄ, T.: The epidemiological challenge of the most frequent eye cancer: retinoblastoma, an issue of birth and death. In: *Br J Ophthalmol* (2009), Nr. 93, S. 1129–1131. <http://dx.doi.org/10.1136/bjo.2008.150292>. – DOI 10.1136/bjo.2008.150292
- [Kohli u. Zhang 2013] KOHLI, RM. ; ZHANG, Y.: TET enzymes, TDG and the dynamics of DNA demethylation. In: *Nature* 502 (2013), Nr. 7472, S. 472–479
- [Krokan u. Bjørås 2013] KROKAN, HE. ; BJØRÅS, M.: Base excision repair. In: *Cold Spring Harbor perspectives in biology* 5 (2013), Nr. 4, S. a012583
- [Laired 2010] LAIRED, PW.: Principles and challenges of genomewide DNA methylation analysis. In: *Nat Rev Genet* 3 (2010), Nr. 11, S. 191–203. <http://dx.doi.org/10.1038/nrg2732>. – DOI 10.1038/nrg2732
- [Landers u. a. 2005] LANDERS, M. ; CALCIANO, MA. ; COLOSI, D. ; GLATT-DEELEY, H. ; WAGSTAFF, J. ; LALANDE, M.: Maternal disruption of Ube3a leads to increased expression of Ube3a-ATS in trans. In: *Nucleic acids research* 33 (2005), Nr. 13, S. 3976–3984
- [Lee u. a. 1986] LEE, WH. ; SHEW, JY. ; HONG, FD. ; SERY, TW. ; DONOSO, LA. ; YOUNG, LJ. ; BOOKSTEIN, R. ; LEE, EYHP.: The retinoblastoma susceptibility gene encodes a

- nuclear phosphoprotein associated with DNA binding activity. In: *Nature* (1986), Nr. 329, S. 642–645. <http://dx.doi.org/10.1038/329642a0>. – DOI 10.1038/329642a0
- [Li 2002] LI, E.: Chromatin modification and epigenetic reprogramming in mammalian development. In: *Nature Reviews Genetics* 3 (2002), Nr. 9, S. 662–673
- [Li u. a. 1992] LI, E. ; BESTOR, TH. ; JAENISCH, R.: Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. In: *Cell* (1992), Nr. 69, S. 915–926. [http://dx.doi.org/10.1016/0092-8674\(92\)90611-F](http://dx.doi.org/10.1016/0092-8674(92)90611-F). – DOI 10.1016/0092-8674(92)90611-F
- [Li u. Durbin 2010] LI, H. ; DURBIN, R.: Fast and accurate long-read alignment with Burrows-Wheeler transform. In: *Bioinformatics* 26 (2010), Nr. 5, S. 589–595
- [Li u. a. 2009] LI, H. ; HANDSAKER, B. ; WYSOKER, A. ; FENNEL, T. ; RUAN, J. ; HOMER, N. ; MARTH, G. ; ABECASIS, G. ; DURBIN, R. ; 1000GENOMEPROJECTDATAPROCESSINGSUBGROUP: The sequence alignment/map format and SAMtools. In: *Bioinformatics* 25 (2009), Nr. 16, S. 2078–2079
- [Li u. a. 2014] LI, W. ; LIU, D. ; CHANG, W. ; LU, X. ; WANG, YL. ; WANG, H. ; ZHU, C. ; LIN, HY. ; ZHANG, Y. ; ZHOU, J. ; WANG, H.: Role of IGF2BP3 in trophoblast cell invasion and migration. In: *Cell Death and Disease* (2014), Nr. 5. <http://dx.doi.org/10.1038/cddis.2013.545>. – DOI 10.1038/cddis.2013.545
- [Liao u. a. 2011] LIAO, B. ; HU, Y. ; BREWER, G.: RNA-binding protein insulin-like growth factor mRNA-binding protein 3 (IMP-3) promotes cell survival via insulin-like growth factor II signaling after ionizing radiation. In: *Journal of Biological Chemistry* 286 (2011), Nr. 36, S. 31145–31152
- [Liao u. a. 2005] LIAO, B. ; HU, Y. ; HERRICK, DJ. ; BREWER, G.: The RNA-binding protein IMP-3 is a translational activator of insulin-like growth factor II leader-3 mRNA during proliferation of human K562 leukemia cells. In: *Journal of Biological Chemistry* 280 (2005), Nr. 18, S. 18517–18524
- [Lister u. a. 2010] LISTER, R. ; PELIZZOLA, M. ; DOWEN, RH. ; HAWKINS, RD. ; HON, G. ; TONTI-FILIPPINI, J. ; NERY, JR. ; LEE, L. ; YE, Z. ; NGO, QM. ; EDSALL, L. ; ANTOSIEWICZ-BOURGET, J. ; STEWART, R. ; RUOTTI, V. ; MILLAR, AH. ; THOMSON, JA. ; REN, B. ; ECKER, JR.: Human DNA methylomes at base resolution show widespread epigenomic differences. In: *Nature* (2010), Nr. 462, S. 315–322. <http://dx.doi.org/10.1038/nature08514>. – DOI 10.1038/nature08514
- [Luedi u. a. 2005] LUEDI, PP. ; HARTEMINK, AJ. ; JIRTLE, RL.: Genome-wide prediction of imprinted murine genes. In: *Genome Res* 6 (2005), Nr. 15, S. 874–884. <http://dx.doi.org/10.1101/gr.3303505>. – DOI 10.1101/gr.3303505
- [Macdonald u. Mann 2014] MACDONALD, WA. ; MANN, MRW.: Epigenetic regulation of genomic imprinting from germ line to preimplantation. In: *Mol Reprod Dev* 81 (2014), S. 126–140. <http://dx.doi.org/10.1002/mrd.22220>. – DOI 10.1002/mrd.22220

- [Maher u. Reik 2000] MAHER, ER. ; REIK, W.: Beckwith-Wiedemann syndrome: imprinting in clusters revisited. In: *The Journal of clinical investigation* 105 (2000), Nr. 3, S. 247–252
- [Martin 2011] MARTIN, M.: Cutadapt removes adapter sequences from highthroughput sequencing reads. In: *EMBnet J* 17 (2011), Nr. 1, S. 10–12
- [Mattick u. a. 2009] MATTICK, JS. ; AMARAL, PP. ; DINGER, ME. ; MERCER, TR. ; MEHLER, MF.: RNA regulation of epigenetic processes. In: *Bioessays* 31 (2009), Nr. 1, S. 51–59
- [McCole u. a. 2011] MCCOLE, RB. ; LOUGHRAN, NB. ; CHAHAL, M. ; FERNANDES, LP. ; ROBERTS, RG. ; FRATERNALI, F. ; O'CONNELL, MJ. ; O'KEY, RJ.: A case-by-case evolutionary analysis of four imprinted retrogenes. In: *Evolution* 65 (2011), Nr. 5, S. 1413–1427. <http://dx.doi.org/10.1111/j.1558-5646.2010.01213.x> – DOI 10.1111/j.1558-5646.2010.01213.x
- [McGrath u. Solter 1984] MCGRATH, J. ; SOLTER, D.: Completion of mouse embryogenesis requires both the maternal and paternal genomes. In: *Cell* 37 (1984), Nr. 1, S. 179–183
- [Meng u. a. 2012] MENG, L. ; PERSON, RE. ; BEAUDET, AL.: Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. In: *Human molecular genetics* 21 (2012), Nr. 13, S. 3001–3012
- [Mighell u. a. 2000] MIGHELL, AJ. ; SMITH, NR. ; ROBINSON, PA. ; MARKHAM, AF.: Vertebrate pseudogenes. In: *FEBS Letters* (2000), Nr. 468, S. 109–114. [http://dx.doi.org/10.1016/S0014-5793\(00\)01199-6](http://dx.doi.org/10.1016/S0014-5793(00)01199-6) – DOI 10.1016/S0014-5793(00)01199-6
- [Monk u. a. 2002] MONK, D. ; BENTLEY, L. ; BEECHEY, C. ; HITCHINS, M. ; PETERS, J. ; PREECE, MA. ; STANIER, P. ; MOORE, GE.: Characterisation of the growth regulating gene IMP3, a candidate for Silver-Russell syndrome. In: *Journal of medical genetics* 39 (2002), Nr. 8, S. 575–581
- [Moore u. Haig 1991] MOORE, T. ; HAIG, D.: Genomic imprinting in mammalian development: a parental tug-of-war. In: *Trends in Genetics* 7 (1991), Nr. 2, S. 45–49
- [Mullis u. a. 1986] MULLIS, K. ; FALOONA, F. ; SCHARF, S. ; R. SAIKI ; HORN, G. ; ERLICH, H.: Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In: *Cold Spring Harb Symp Quant Biol* 51 (1986), Nr. Pt 1, S. 263–273
- [Nakabayashi u. a. 2011] NAKABAYASHI, K. ; TRUJILLO, AM. ; TAYAMA, C. ; CAMPRUBI, C. ; YOSHIDA, W. ; LAPUNZINA, P. ; SANCHEZ, A. ; SOEJIMA, H. ; ABURATANI, H. ; NAGAE, G. ; OGATA, T. ; HATA, K. ; MONK, D.: Methylation screening of reciprocal genome-wide UPDs identifies novel human-specific imprinted genes. In: *Hum Mol Genet* 20 (2011), Nr. 16, S. 3188–3197. <http://dx.doi.org/10.1093/hmg/ddr224> – DOI 10.1093/hmg/ddr224
- [Neumann u. a. 2014] NEUMANN, LC. ; FEINER, N. ; MEYER, A. ; BUITING, K. ; HORSTHEMKE, B.: The Imprinted NPAP1 Gene in the Prader–Willi Syndrome Region

- Belongs to a POM121-Related Family of Retrogenes. In: *Genome biology and evolution* 6 (2014), Nr. 2, S. 344–351
- [Nordheim u. Knippers 2015] NORDHEIM, A. ; KNIPPERS, R.: *Molekulare Genetik*. 10. Georg Thieme Verlag, 2015. – ISBN 978-3-13-477010-0
- [Ohno 1970] OHNO, S.: Evolution by gene duplication. In: *Evolution by gene duplication*. (1970)
- [Ohshima u. a. 2003] OHSHIMA, K. ; HATTORI, M. ; YADA, T. ; GOJOBORI, T. ; SAKAKI, Y. ; OKADA, N.: Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repetas by particular LI subfamilies in ancestral primates. In: *Genome Biology* 4 (2003). <http://dx.doi.org/10.1038/nature08514>. – DOI 10.1038/nature08514
- [Okae u. a. 2014] OKAE, H. ; CHIBA, H. ; HIURA, H. ; HAMADA, H. ; SATO, A. ; UTSUNOMIYA, T. ; KIKUCHI, H. ; YOSHIDA, H. ; TANAKA, A. ; SUYAMA, M. ; ARIMA, T.: Genome-wide analysis of DNA methylation dynamics during early human development. In: *PLoS genetics* 10 (2014), Nr. 12, S. e1004868
- [Okano u. a. 1999] OKANO, M. ; BELL, DW. ; HABER, DA. ; LI, E.: DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. In: *Cell* (1999), Nr. 99, S. 247–257. [http://dx.doi.org/10.1016/S0092-8674\(00\)81656-6](http://dx.doi.org/10.1016/S0092-8674(00)81656-6). – DOI 10.1016/S0092-8674(00)81656-6
- [Pastor u. a. 2013] PASTOR, WA. ; ARAVIND, L. ; RAO, A.: TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. In: *Nature reviews Molecular cell biology* 14 (2013), Nr. 6, S. 341–356
- [Perelman u. a. 2011] PERELMAN, P. ; JOHNSON, WE. ; ROOS, C. ; SEUANEZ, HN. ; HORVATH, JE. ; MOREIRA, MAM. ; KESSING, B. ; PONTIUS, J. ; ROELKE, M. ; RUMPLER, Y. ; SCHNEIDER, MPC. ; SILVA, A. ; O'BRIEN, SJ. ; PESON-SLATTERY, J.: A molecular phylogeny of living primates. In: *PLoS Genet* 7 (2011), Nr. 3. <http://dx.doi.org/10.1371/journal.pgen.1001342>. – DOI 10.1371/journal.pgen.1001342
- [Prickett u. Oakey 2012] PRICKETT, AR. ; OAKEY, RJ.: A survey of tissue-specific genomic imprinting in mammals. In: *Molecular genetics and genomics* 287 (2012), Nr. 8, S. 621–630
- [Pruitt u. a. 2009] PRUITT, KD. ; HARROW, J. ; HARTE, RA. ; WALLIN, C. ; DIEKHANS, M. ; MAGLOTT, DR. ; SEARLE, S. ; FARRELL, CM. ; LOVELAND, JE. ; RUEF, BJ. ; HART, E. ; SUNER, M. ; LANDRUM, MJ. ; AKEN, B. ; AYLING, S. ; BAERTSCH, R. ; FERNANDEZ-BANET, J. ; CHERRY, JL. ; CURWEN, V. ; DiCUCCIO, M. ; KELLIS, M. ; LEE, J. ; LIN, MF. ; SCHUSTER, M. ; SHKEDA, A. ; AMID, C. ; BROWN, G. ; DUKHANINA, O. ; RANKISH, A. ; HART, J. ; MAIDAK, BL. ; MUDGE, J. ; MURPHY, MR. ; MURPHY, T. ; RAJAN, J. ; RAJPUT, B. ; RIDDICK, LD. ; SNOW, C. ; WU, W. ; BIRNEY, E. ; HAUSSLER, D. ; HUBBARD, T. ; OSTELL, J. ; DURBIN, R. ; LIPMAN, D.: The consensus coding sequence (CCDS) project: Identifying a common protein-coding

- gene set for the human mouse genomes. In: *Genome Res* 19 (2009), Nr. 7, S. 1316–1323. <http://dx.doi.org/10.1101/gr.080531.108>. – DOI 10.1101/gr.080531.108
- [Rademacher u. a. 2014] RADEMACHER, K. ; SCHRÖDER, C. ; KANBER, D. ; KLEIN-HITPASS, L. ; WALLNER, S. ; ZESCHNIGK, M. ; HORSTHEMKE, B.: Evolutionary origin and methylation status of human intronic CpG islands that are not present in mouse. In: *Genome Biol Evol* 6 (2014), Nr. 7, S. 1579–1588. <http://dx.doi.org/doi.1093/gbe/evu125>. – DOI doi.1093/gbe/evu125
- [Rahmann u. a. 2013] RAHMANN, S. ; BEYGO, J. ; KANBER, D. ; MARTIN, M. ; HORSTHEMKE, B. ; BUITING, K.: Amplifyer: automated methylation analysis of amplicons from bisulfite flowgram sequencing. In: *PeerJ PrePrints* (2013), Nr. 1. <http://dx.doi.org/10.7287/peerj.preprints.122v2>. – DOI 10.7287/peerj.preprints.122v2
- [Rapkins u. a. 2006] RAPKINS, RW. ; HORE, T. ; SMITHWICK, M. ; AGER, E. ; PASK, AJ. ; RENFREE, MB. ; KOHN, M. ; HAMEISTER, H. ; NICHOLLS, RD. ; DEAKIN, JE. ; GRAVES, JAM.: Recent assembly of an imprinted domain from non-imprinted components. In: *PLoS Genet* 2 (2006), Nr. 10, S. e182
- [Rees u. a. 1993] REES, WA. ; YAGER, TD. ; KORTE, J. ; HIPPEL, PH. von: Betaine can eliminate the base pair composition dependence of DNA melting. In: *Biochemistry* 32 (1993), Nr. 1, S. 137–144
- [Reik u. a. 2001] REIK, W. ; DEAN, W. ; WALTER, J.: Epigenetic reprogramming in mammalian development. In: *Science* 293 (2001), Nr. 5532, S. 1089–1093
- [Reik u. Walter 2001] REIK, W. ; WALTER, J.: Genomic imprinting: parental influence on the genome. In: *Nature Reviews Genetics* 2 (2001), Nr. 1, S. 21–32
- [Robertson 2005] ROBERTSON, KD.: DNA methylation and human disease. In: *Nature Reviews Genetics* 6 (2005), Nr. 8, S. 597–610
- [Rougeulle u. a. 1997] ROUGEULLE, C. ; GLATT, H. ; LALANDE, M.: The Angelman syndrome candidate gene, UBE3A/MECP2, is imprinted in brain. In: *Nature genetics* 17 (1997), Nr. 1, S. 14–15
- [Royo u. Cavaillé 2008] ROYO, H. ; CAVAILLÉ, J.: Non-coding RNAs in imprinted gene clusters. In: *Biology of the Cell* 100 (2008), Nr. 3, S. 149–166
- [Sanger u. a. 1977] SANGER, F. ; NICKLEN, S. ; COULSON, AR: DNA sequencing with chain-terminating inhibitors. In: *Proc Natl Acad Sci U S A* 74 (1977), Nr. 12, S. 5463–5467
- [Saxonov u. a. 2006] SAXONOV, S. ; BERG, P. ; BRUTLAG, DL.: A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. In: *Proc Natl Acad Sci USA* 103 (2006), Nr. 5, S. 1412–1417. <http://dx.doi.org/10.1073/pnas.0510310103>. – DOI 10.1073/pnas.0510310103
- [Schübeler 2015] SCHÜBELER, D.: Function and information content of DNA methylation. In: *Nature* 517 (2015), Nr. 7534, S. 321–326

- [Seisenberger u. a. 2013] SEISENBERGER, S. ; PEAT, JR. ; HORE, TA. ; SANTOS, F. ; DEAN, W. ; REIK, W.: Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. In: *Phil Trans R Soc B* 386 (2013). <http://dx.doi.org/10.1098/rstb.2011.0330>. – DOI 10.1098/rstb.2011.0330
- [Shearwin u. a. 2005] SHEARWIN, KE. ; CALLEN, BBP. ; EGAN, JB.: Transcriptional interference—a crash course. In: *TRENDS in Genetics* 21 (2005), Nr. 6, S. 339–345
- [Stadler u. a. 2011] STADLER, MB. ; MURR, R. ; BURGER, L. ; IVANEK, R. ; LIENERT, F. ; SCHÖLER, A. ; NIMWEGEN, E. van ; WIRBELAUER, C. ; OAKELEY, EJ. ; GAIDATZIS, D. ; TIWARI, VK. ; SCHÜBELER, D.: DNA-binding factors shape the mouse methylome at distal regulatory regions. In: *Nature* (2011)
- [Steenpass u. a. 2013] STEENPASS, L. ; KANBER, D. ; HIBER, M. ; BUITING, K. ; HORSTHEMKE, B. ; LOHMANN, D.: Human PPP1R26P1 functions as cis-repressive element in mouse Rb1. In: *PLoS One* 8 (2013), Nr. 9. <http://dx.doi.org/10.1371/journal.pone.0074159>. – DOI 10.1371/journal.pone.0074159
- [Surani u. a. 1984] SURANI, MAH. ; BARTON, SC. ; NORRIS, ML.: Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. In: *Nature* 308 (1984), Nr. 5959, S. 548–550
- [Suvasini u. a. 2011] SUVASINI, R. ; SHRUTI, B. ; THOTA, B. ; SHINDE, SV. ; FRIEDMANN-MORVINSKI, D. ; NAWAZ, Z. ; PRASANNA, KV. ; THENNARASU, K. ; HEGDE, AS. ; ARIVAZHAGAN, A. ; CHANDRAMOULI, BA. ; SANTOSH, V. ; SOMASUNDARAM, K.: Insulin growth factor-2 binding protein 3 (IGF2BP3) is a glioblastoma-specific marker that activates phosphatidylinositol 3-kinase/mitogen-activated protein kinase (PI3K/MAPK) pathways by modulating IGF-2. In: *Journal of Biological Chemistry* 286 (2011), Nr. 29, S. 25882–25890
- [Suzuki u. a. 2007] SUZUKI, S. ; ONO, R. ; NARITA, T. ; PASK, AJ. ; SHAW, G. ; WANG, C. ; KOHDA, T. ; ALSOP, AE. ; GRAVES, JA. M. ; KOHARA, Y. ; ISHINO, F. ; RENFREE, MB. ; KANEKO-ISHINO, T.: Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. In: *PLoS Genet* 3 (2007), Nr. 4, S. e55
- [Suzuki u. a. 2011] SUZUKI, S. ; SHAW, G. ; KANEKO-ISHINO, T. ; ISHINO, F. ; RENFREE, MB.: The evolution of mammalian genomic imprinting was accompanied by the acquisition of novel CpG islands. In: *Genome Biol Evol* 3 (2011), S. 1276–1283. <http://dx.doi.org/10.1038/nature08514>. – DOI 10.1038/nature08514
- [Takai u. Jones 2002] TAKAI, D. ; JONES, PA.: Comprehensive analysis of CpG islands in human chromosomes 21 and 22. In: *PNAS* 99 (2002), Nr. 6, S. 3740–3745. <http://dx.doi.org/10.1073/pnas.052410099>. – DOI 10.1073/pnas.052410099
- [Thorvaldsdottir u. a. 2013] THORVALDSDOTTIR, H. ; ROBINSON, JT. ; MESIROV, JP.: Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. In: *Brief Bioinform.* 14 (2013), Nr. 2, S. 178–192

- [Torrents u. a. 2003] TORRENTS, D. ; SUYAMA, M. ; ZDOBNOV, E. ; BORK, P.: A genome-wide survey of human pseudogenes. In: *Genome research* 13 (2003), Nr. 12, S. 2559–2567
- [Tweedie u. a. 1997] TWEEDIE, S. ; CHARLTON, J. ; CLARK, V. ; BIRD, A.: Methylation of genomes and genes at the invertebrate-vertebrate boundary. In: *Molecular and cellular biology* 17 (1997), Nr. 3, S. 1469–1475
- [Vinckenbosch u. a. 2006] VINCKENBOSCH, N. ; DUPANLOUP, I. ; KAESMANN, H.: Evolutionary fate of retroposed gene copies in the human genome. In: *PNAS* 103 (2006), Nr. 9, S. 3220–3225. <http://dx.doi.org/10.1073/pnas.0511307103>. – DOI 10.1073/pnas.0511307103
- [Wei u. a. 2001] WEI, W. ; GILBERT, N. ; OOI, SL. ; LAWLER, JF. ; OSTERTAG, EM. ; KAZAZIAN, HH. ; BOEKE, JD. ; MORAN, JV.: Human L1 retrotransposition cis preference versus trans complementation. In: *Mol Cell Biol* 21 (2001), Nr. 4, S. 1429–1439. <http://dx.doi.org/10.1128/MCB.21.4.1429-1439.2001>. – DOI 10.1128/MCB.21.4.1429–1439.2001
- [Wood u. a. 2007] WOOD, AJ. ; ROBERTS, RG. ; MONK, D. ; MOORE, GE. ; SCHULZ, R. ; OAKEY, RJ.: A screen for retroposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. In: *PLoS Genet* 3 (2007), Nr. 2. <http://dx.doi.org/10.1371/journal.pgen.0030020>. – DOI 10.1371/journal.pgen.0030020
- [Wu u. Zhang 2014] WU, H. ; ZHANG, Y.: Reversing DNA methylation: mechanisms, genomics, and biological functions. In: *Cell* 156 (2014), Nr. 1, S. 45–68
- [Zhang u. a. 2011] ZHANG, A. ; SKAAR, DA. ; LI, Y. ; HUANG, D. ; PRICE, TM. ; MURPHY, SK ; JIRTLE, RL.: Novel retrotransposed imprinted locus identified at human 6p25. In: *Nuc Acids Res* 39 (2011), Nr. 13, S. 5344–5400. <http://dx.doi.org/10.1093/nar/gkr108>. – DOI 10.1093/nar/gkr108
- [Zhang 2003] ZHANG, J.: Evolution by gene duplication: an update. In: *TRENDS in Ecology and Evolution* (2003), Nr. 6, S. 292–298. [http://dx.doi.org/10.1016/S0169-5347\(03\)00033-8](http://dx.doi.org/10.1016/S0169-5347(03)00033-8). – DOI 10.1016/S0169–5347(03)00033–8
- [Zhang u. a. 2000] ZHANG, Z. ; SCHWARTZ, S. ; WAGNER, L. ; MILLER, W.: A greedy algorithm for aligning DNA sequences. In: *J Comput Biol* 7 (2000), Nr. 1-2, S. 203–214
- [Zhou u. a. 2011] ZHOU, VW. ; GOREN, A. ; BERNSTEIN, BE.: Charting histone modifications and the functional organization of mammalian genomes. In: *Nature Reviews Genetics* 12 (2011), Nr. 1, S. 7–18
- [Ziller u. a. 2013] ZILLER, MJ. ; GU, H. ; MÜLLER, F. ; DONAGHEY, J. ; TSAI, LTY. ; KOHLBACHER, O. ; JAGER, PL. D. ; ROSEN, ED. ; BENNETT, DQ. ; BERNSTEIN, BE. ; GNIRKE, A. ; MEISSNER, A.: Charting a dynamic DNA methylation landscape of the human genome. In: *Nature* 500 (2013), Nr. 7463. <http://dx.doi.org/10.1038/nature12433>. – DOI 10.1038./nature12433

Abbildungsverzeichnis

1.1	Schematische Darstellung der <i>RB1</i> -Gens im Humangenom	11
2.1	Schematische Darstellung von möglichen Positionen eines CGIs in Abhängigkeit von einem CCDS	20
2.2	Schematische Darstellung und Berechnung der Distanzen von Lokalisierungs- und Nachbarschaftsdatei	22
2.3	Schematische Darstellung der zwei aufeinanderfolgenden PCRs bei der Herstellung von <i>Amplicon libraries</i> . Die Graphik wurde der Roche <i>Sequencing Application Note</i> entnommen (Horsthemke u. a., 2013).	35
3.1	Methylierungsmuster von drei human/nicht-murin intronischen CGIs mit einem Methylierungsgrad zwischen 20% und 80%	46
3.2	Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozygoten Individuen	50
3.3	Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Blut von heterozygoten Individuen, die informativ in Bezug auf die elterliche Herkunft sind	51
3.4	Entstehungszeitpunkte der fünf human/nicht-murin intronischen CGIs mit einem mittleren Methylierungsgrad	53
3.5	Schematische Darstellung der Introns, die ein CGI mit einem mittleren Methylierungsgrad beinhalten	54
3.6	Flußdiagramm der genomweiten Analyse von Retrokopie-assoziierten CGIs	55
3.7	Grad der Methylierung der Retrokopie-assoziierten CGIs in Abhängigkeit der Anzahl ihrer Vorkommen	58
3.8	Clusteranalyse vom Methylierungsgrad der R-CGIs in Abhängigkeit vom Methylierungsgrad der Umgebung der Retrokopie	59
3.9	Schematische Darstellung der Lage des R-CGIs 9261_1_hg19 im humanen Genom	65
3.10	Histonmodifikationen des R-CGIs 9261_1_hg19	66
3.11	Methylierungsanalyse des R-CGIs 9261_1_hg19 in Blut von drei heterozygoten Individuen	67
3.12	Methylierungsanalyse des R-CGIs 9261_1_hg19 in drei Plazenta-Proben	68
3.13	Methylierungsanalyse des R-CGIs 9261_1_hg19 von sechs Patienten mit einem Multilocus- <i>Imprinting</i> -Defekt und vier Normalkontrollen	69

3.14	Expressionsanalyse des R-CGIs 9261_1_hg19 (Blut)	70
3.15	Methylierungsanalyse der beiden CGIs (CpG87 und CpG144) des <i>IGF2BP3</i> - Gens	71
3.16	Expressionsanalyse von <i>IGF2BP3</i> in Blut	72
3.17	Expressionsanalyse von <i>IGF2BP3</i> in Plazentaprobe aus unterschiedlichen Trimestern	72
A.1	Histogramme über die Methylierung einzelner Reads für human/nicht-murin intronische CGIs mit einem mittleren Methylierungsniveau - Part 1	108
A.2	Histogramme über die Methylierung einzelner Reads für human/nicht-murin intronische CGIs mit einem mittleren Methylierungsniveau - Part 2	109
A.3	Histogramme über die Methylierung einzelner Reads für human/nicht-murin intronische CGIs mit einem mittleren Methylierungsniveau - Part 3	110
A.4	Histogramme über die Methylierung einzelner Reads für human/nicht-murin intronische CGIs mit einem mittleren Methylierungsniveau - Part 4	111
A.5	Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozyoten Individuen - Part 1	113
A.6	Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozyoten Individuen - Part 2	114
A.7	Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozyoten Individuen - Part 3	115
A.8	Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozyoten Individuen - Part 4	116
A.9	Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozyoten Individuen - Part 5	117
A.10	Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozyoten Individuen - Part 6	118
A.11	Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozyoten Individuen - Part 7	119
A.12	Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Blut von hetero- zygoten Individuen, die teilweise informativ in Bezug auf die elterliche Herkunft sind	120
A.13	Methylierungsanalyse einzelner CpGs der 12 R-CGIs in Monozyten von heterozyoten Individuen - Part 1	126
A.14	Methylierungsanalyse einzelner CpGs der 12 R-CGIs in Monozyten von heterozyoten Individuen - Part 2	127

A.15 Methylierungsanalyse einzelner CpGs der 12 R-CGIs in Monozyten von heterozyten Individuen - Part 3	128
A.16 Methylierungsanalyse einzelner CpGs der 12 R-CGIs in Monozyten von heterozyten Individuen - Part 4	129
A.17 Methylierungsplot vom Methylierungsgrad der R-CGIs in Abhängigkeit vom Methylierungsgrad der Umgebung der Retrokopie, in der sowohl 1000 bp <i>upstream</i> als auch <i>downstream</i> repetitive Sequenzen liegen	134
A.18 Methylierungsplot vom Methylierungsgrad der R-CGIs in Abhängigkeit vom Methylierungsgrad der Umgebung der Retrokopie, in der keine repetitive Sequenzen liegen	135
A.19 Methylierungsplot vom Methylierungsgrad der R-CGIs in Abhängigkeit vom Methylierungsgrad der Umgebung der Retrokopie, wobei 1000 bp <i>upstream</i> der Retrokopie repetitive Sequenzen liegen	136
A.20 Methylierungsplot vom Methylierungsgrad der R-CGIs in Abhängigkeit vom Methylierungsgrad der Umgebung der Retrokopie, wobei 1000 bp <i>downstream</i> der Retrokopie repetitive Sequenzen liegen	137

Tabellenverzeichnis

3.1	Lage aller humanen und murinen CGIs in Abhängigkeit von proteinkodierenden Regionen (CCDS)	42
3.2	Methylierungsanalyse von 104 human/nicht-murin intronischen CGIs . . .	44
3.3	Methylierungsanalysen einzelner Reads und CpG-Dinukleotide der CGIs mit einem mittleren Methylierungslevel (Methylom 1 und Methylom 2) . .	45
3.4	Methylierungsgrad der fünf potentiell differentiell methylierten CGIs in anderen Geweben	48
3.5	Evolutionäre Analyse von 86 human/nicht-murin intronischen CGIs in unterschiedlichen Primaten	52
3.6	Häufigkeit von Retrokopie-assoziierten CGIs	56
3.7	Methylierungsanalyse einzelner Reads und CpG-Dinukleotide der Retrokopie-assoziierten CGIs mit einem mittleren Methylierungsgrad (Methylom 1) . .	62
3.8	Methylierungsanalyse der zwölf R-CGIs mittels Roche/454 GS Junior . . .	63
A.1	Genomische Primersequenzen. Teile dieser Tabelle wurden schon in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.	100
A.2	Bisulfit Primersequenzen, universelle <i>tag</i> -Sequenzen in fett dargestellt. Teile dieser Tabelle wurden schon in abgeänderter Form in Rademacher u. a. (2014) und Grothaus u. a. (2016) veröffentlicht.	101
A.3	cDNA Primersequenzen für <i>RPS2P32</i> und <i>IGF2BP3</i> . Diese Tabelle wurden schon in abgeänderter Form in Grothaus u. a. (2016) veröffentlicht.	101
A.4	SNaPshot Primersequenzen für <i>RPS2P32</i> (SNP: rs10228640) und <i>IGF2BP3</i> (SNP: rs156414). Diese Tabelle wurden schon in abgeänderter Form in Grothaus u. a. (2016) veröffentlicht.	102
A.5	Sequenzen der MID-Primer	102
A.5	Sequenzen der MID-Primer	103
A.6	Genotypisierung von 14414_1_hg19(<i>ASRGL1</i>), 15224_1_hg19 (<i>PARP11</i>) und 19100_1_hg19 (<i>PDXDC1</i>) in 22 Monozyten-Proben	104
A.7	Genotypisierung von 2391_1_hg19, 6141_1_hg19, 9224_1_hg19, 9261_1_hg19, 9377_1_hg19 und 9473_1_hg19 in 22 Monozyten-Proben	104
A.8	Genotypisierung von 13250_1_hg19, 15400_1_hg19, 15512_1_hg19, 16448_1_hg19 und 16458_1_hg19 in 22 Monozyten-Proben	105
A.9	Genotypisierung von 17031_1_hg19, 20403_1_hg19, 23548_1_hg19, 24982_1_hg19 und 26859_1_hg19 in 22 Monozyten-Proben	105

A.10 Methylierungsdaten von 104 human/nicht-murinen intronischen CGIs mit zusätzlichen Treffern im Genom (Methylom 1 und 2). Diese Tabelle wurde schon in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht. . . .	106
A.10 Methylierungsdaten von 104 human/nicht-murinen intronischen CGIs mit zusätzlichen Treffern im Genom (Methylom 1 und 2). Diese Tabelle wurde schon in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht. . . .	107
A.11 Methylierungsgrad des CGIs 14414_1_hg19(<i>ASRGL1</i>) in Monozyten . . .	112
A.12 Methylierungsgrad des CGIs 15224_1_hg19(<i>PARP11</i>) in Monozyten . .	112
A.13 Methylierungsgrad des CGIs 19100_1_hg19(<i>PDXDC1</i>) in Monozyten . .	112
A.14 Methylierungsgrad der CGIs 14414_1_hg19(<i>ASRGL1</i>), 15224_1_hg19(<i>PARP11</i>) und 19100_1_hg19(<i>PDXDC1</i>) in Blut	119
A.15 Expressionslevel aller R-CGIs in Abhängigkeit ihrer Methylierung	121
A.16 Genomische Lage (RefSeq) aller R-CGIs in Abhängigkeit ihrer Methylierung	121
A.17 Genomische Lage (RefSeq) aller R-CGIs in Abhängigkeit ihres Expressionslevels	121
A.18 Expressionslevel aller <i>intergenic</i> R-CGIs in Abhängigkeit ihrer Methylierung	122
A.19 Expressionslevel aller <i>intragenic</i> R-CGIs in Abhängigkeit ihrer Methylierung	122
A.20 Methylierungsgrad von zwei Retrokopie-assoziierten CGIs in anderen Gewebetypen	123
A.21 Methylierungsgrad von zehn Retrokopie-assoziierten CGIs in anderen Gewebetypen	124
A.22 Methylierungsgrad von fünf Retrogen-überlappenden CGIs in anderen Geweben	125
A.23 Evolutionsanalyse von 104 human/nicht-murinen intronischen CGIs mit zusätzlichen Treffern im Genom. (<i>gap</i>) Sequenzlücke im Genom, * Teile des Gens nicht vorhanden. Diese Tabelle wurde schon in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.	130
A.23 Evolutionsanalyse von 104 human/nicht-murinen intronischen CGIs mit zusätzlichen Treffern im Genom. (<i>gap</i>) Sequenzlücke im Genom, * Teile des Gens nicht vorhanden. Diese Tabelle wurde schon in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.	131
A.24 RT-PCR Ansätze von <i>RPS2P32</i> , <i>IGF2BP3</i> und β - <i>Aktin</i>	132
A.25 Rohdaten der quantitativen Analyse des allelischen Transkriptverhältnisses zu <i>RPS2P32</i>	138
A.26 Rohdaten der quantitativen Analyse des allelischen Transkriptverhältnisses des <i>IGF2BP3</i> -Gens	138

A Anhang Dissertation

A.1 URLs von benutzen Datenbanken und Programmen

Alle URLs zu benutzen Datenbanken und Programmen:

UCSC Browser	https://genome.ucsc.edu/
HUGO	http://www.genenames.org/
NCBI CCDS Datenbank	http://www.ncbi.nlm.nih.gov/CCDS/
ENA	http://www.ebi.ac.uk/ena
EGA	https://www.ebi.ac.uk/ega/
GEO NCBI	http://www.ncbi.nlm.nih.gov/geo/
JGA	http://trace.ddbj.nig.ac.jp/jga/index_e.html
Perl	http://www.perl.org/
R	http://www.r-project.org/
blast2seq	http://blast.ncbi.nlm.nih.gov/Blast.cgi
MEGA-BLAST	http://www.ncbi.nlm.nih.gov/blast/mmtrace.shtml
Primer3	http://primer3.ut.ee/
MethPrimer	http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi
UCSC <i>in silico</i> PCR	https://genome.ucsc.edu/cgi-bin/hgPcr?command=start
BiSearch	http://bisearch.enzim.hu/
BLAT	https://genome.ucsc.edu/cgi-bin/hgBlat?command=start
Sequencing Analysis	http://www.appliedbiosystems.com/absite/us/en/home/support/software/dna-sequencing/sequencing-analysis.html
Geneious	http://www.geneious.com/
cutadapt	https://code.google.com/p/cutadapt/
methyIcTools	http://www.dkfz.de/en/genetics/staff/staff_detail/Hovestadt.html
BWA	http://bio-bwa.sourceforge.net/
SAMtools	http://samtools.sourceforge.net/
MarkDuplicates	http://picard.sourceforge.net
Skript bedGraphToBigWig	http://genome.ucsc.edu/
RepeatMasker	http://repeatmasker.org

A.2 Primersequenzen

A.2.1 Genomische Primersequenzen

Tabelle A.1: Genomische Primersequenzen. Teile dieser Tabelle wurden schon in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

Primername	Primersequenz (5'-3')	Tm (°C)	Produktgröße (bp)	Zyklen	Betain
ASRGL1-CpG28-F ASRGL1-CpG28-R3	TCAAAAGCTCCCTTTTGCAC CTGGTTTCTCATCACTCCAA	64	516	45	ja
PARP11-CpG88-F2 PARP11-CpG88-R	CTGGCCATGTGGGGCTAG ACCCATTCCCTGTGGATTTTC	64	353	35	nein
PDXDC1-CpG69-F PDXDC1-CpG69-R	TGCAGGAAGGGCACAAAGTC CGATGTGGCTACAGCAGG	68	382	45	ja
2391_1_hg19_F 2391_1_hg19_R	GTCTTGCACCAGTCACTCTTG TACTCTACGGCCCTTCACAG	64	272	35	nein
6141_1_hg19_F 6141_1_hg19_R	CTGGAATTTTGGACCATGC CGCTGGGCTTTCTCAAG	64	428	45	ja
9224_1_hg19_F 9224_1_hg19_R	TTAAAGCAAAATCCCAGCCCC AAAAAACAAAGGCGGCTCGC	68	186	35	nein
9261_1_hg19_F 9261_1_hg19_R	GACCAAGCTCTCCATTGTCA GGGATCAGGTAGCTGTAGATG	63	289	35	nein
9377_1_hg19_F 9377_1_hg19_R	GGCTGGCTCTGGTTAGGTC AGCTCAAAGATGCCAGTGAAG	64	318	35	nein
9473_1_hg19_F 9473_1_hg19_R	ACCGCTTTGAAACCCCTACT GGACTTGGGCCTTTGTGTT	62	244	35	nein
13085_1_hg19_F 13085_1_hg19_R	CCACCAAAGGGAAGCTTGTA GCTCTGGGCCATGATCTCT	64	319	35	nein
13250_1_hg19_F 13250_1_hg19_R	CTCTGAACGCCCTGTTCCTTC AGCAGCCTCTCCACTTGTT	64	250	35	nein
15400_1_hg19_F 15400_1_hg19_R	ATTGGGCTCAGATCAGGAG GAAGAAAAGCCTTTCCCAAG	66	650	35	nein
15512_1_hg19_F 15512_1_hg19_R	CCTTCTCACGCAAAGCTTCA GCCCTAATCTGCCTCCTC	68	218	35	nein
16448_1_hg19_F 16448_1_hg19_R	TCCAACACATACTATACCACAG GATGTTCTCGAGCTCCATGG	66	234	45	ja
16458_1_hg19_F 16458_1_hg19_R	GCGTCTGCCCAACTCTTG CTGCCTGCCTGATATGACTG	64	225	35	nein
17031_1_hg19_F 17031_1_hg19_R	GCCCTCCACAAGGAAACT AAAGGCAGATGGCTCATAAC	64	512	45	ja
20403_1_hg19_F 20403_1_hg19_R	GGCGGYGCTCTACTTCTG CCATTGGCTTTCACTCAGAC	64	256	35	nein
23548_1_hg19_F 23548_1_hg19_R	CCATCTGGCTCAAGCTGTC AAGATGATGGGGGAACATGC	68	334	35	nein
24982_1_hg19_F 24982_1_hg19_R	TCATCTTCCGGTTAGCGACA CCAGTTGGCTCTTAATGCCC	64	226	35	nein
26859_1_hg19_F 26859_1_hg19_R	CTAGAGCGTCAGGGATCAGG GGCGACAAACAAGCGTATCT	64	279	35	nein

A.2.2 Bisulfit Primersequenzen

Tabelle A.2: Bisulfit Primersequenzen, universelle *tag*-Sequenzen in fett dargestellt. Teile dieser Tabelle wurden schon in abgeänderter Form in Rademacher u. a. (2014) und Grothaus u. a. (2016) veröffentlicht.

Primername	Primersequenz (5'-3')	Tm (°C)	Produktgröße (bp)
ASRGL1-CpG28-Ftag ASRGL1-CpG28-RM13	CTTGCTTCCTGGCACGAGT TTTTTGTATTTTGTGAGGAAGT CAGGAAACAGCTATGACCT CTAATTTCTCATCACTCCAA	56	543
PARP11-CpG88-Ftag PARP11-CpG88-RM13	CTTGCTTCCTGGCACGAGT TGGTTATGTGGGGTTAGTTTT CAGGAAACAGCTATGACACCC ATTCTATAAAATTTCCA	56	388
PDXDC1-CpG69-Ftag PDXDC1-CpG69-RM13	CTTGCTTCCTGGCACGAGT TGTAYGAAGGGTATAAAGTT CAGGAAACAGCTATGACCC CRATATACTACAACAAAA	50	421
2391_1_hg19_Ftag 2391_1_hg19_RM13	CTTGCTTCCTGGCACGAGT GAGGTTAGTATTTTTTTGATTTG CAGGAAACAGCTATGACCT AAACAACAAAACCAACCATATCTA	56	309
6141_1_hg19_Ftag 6141_1_hg19_RM13	CTTGCTTCCTGGCACGAGT TGAATTGTTTTYGTAGGGTT CAGGAAACAGCTATGACAT CTCTTTTCCATTCCATAATTTA	53	360
9224_1_hg19_Ftag 9224_1_hg19_RM13	CTTGCTTCCTGGCACGAGT GTTGGGGTATATTTTAAAGTAAATTTAG CAGGAAACAGCTATGACAAA ACTTAAACAAATYACAAAAA	54	258
9261_1_hg19_Ftag 9261_1_hg19_RM13	CTTGCTTCCTGGCACGAG ATTGTGATTAAGTTTTTTATTTGTTA CAGGAAACAGCTATGACT CTCCTTCCAAAAATCCTTC	50	350
9473_1_hg19_Ftag 9473_1_hg19_RM13	CTTGCTTCCTGGCACGAGT GGATTGAGGAGTGTTGATAGTTAT CAGGAAACAGCTATGACCC ATAAACTTAAACCTTATATTTTATT	56	307
13250_1_hg19_Ftag 13250_1_hg19_RM13	CTTGCTTCCTGGCACGAGT TTAGATTAGTTTTTGTATTTTGTAG CAGGAAACAGCTATGACAC TCCAACAACCTCTCCACTTAAT	56	233
15400_1_hg19_Ftag 15400_1_hg19_RM13	CTTGCTTCCTGGCACGAGT TGTTGGTATTTTGTAGTTAAATT CAGGAAACAGCTATGACAT AAAAATTATCCAAAAAATTTTCC	50	577
16448_1_hg19_Ftag 16448_1_hg19_RM13	CTTGCTTCCTGGCACGAGT GTTTGTGAGAAATATTATGTT CAGGAAACAGCTATGACAT ATARAAAACTAAACATCTACC	50	342
16458_1_hg19_Ftag 16458_1_hg19_RM13	CTTGCTTCCTGGCACGAGT GATTAGTTTTGTTGTTTGATATGATTG CAGGAAACAGCTATGACAAA ATAATAATAACTCTTTACCATAACC	48	466
17031_1_hg19_Ftag 17031_1_hg19_RM13	CTTGCTTCCTGGCACGAGG TTTAGGGGAGAGGTTTAGG CAGGAAACAGCTATGACCC TAAAACTATTAATCAAATACTCCAAA	56	294
20403_1_hg19_Ftag 20403_1_hg19_RM13	CTTGCTTCCTGGCACGAG ATTTAAGAAGTGTTTTTTTTGTTT CAGGAAACAGCTATGACCT AAATCTTCTACTTCTAAATATCACACT	56	369
23548_1_hg19_Ftag 23548_1_hg19_RM13	CTTGCTTCCTGGCACGAG TTAGTAGGGTTTGTTTTTAGGG CAGGAAACAGCTATGACC ATAAAAAATAAAAAAACTACCCC	56	312

A.2.3 cDNA Primer

Tabelle A.3: cDNA Primersequenzen für *RPS2P32* und *IGF2BP3*. Diese Tabelle wurden schon in abgeänderter Form in Grothaus u. a. (2016) veröffentlicht.

Primername	Primersequenz (5'-3')	Tm (°C)	Produktgröße (bp)
9261_1_hg19_RT_F 9261_1_hg19_RT_R	CAAGCTCTCCATTGTCACC CAAGCTCTCCATTGTCACC	63	289
F_IGF2BP3_cDNA R_IGF2BP3_cDNA	GCTGAGGAGGAGATCATGAAG GAAGTGAAGCTCTGGTGGTC	59	378

A.2.4 SNaPshot Primer

Tabelle A.4: SNaPshot Primersequenzen für *RPS2P32* (SNP: rs10228640) und *IGF2BP3* (SNP: rs156414). Diese Tabelle wurden schon in abgeänderter Form in Grothaus u. a. (2016) veröffentlicht.

Primername	Primersequenz (5'-3')	Tm (°C)	gDNA	cDNA
F_RPS2P32_SNAPSHOT	CACTGCCACCCTGGGCA	65,6	ja	ja
R_RPS2P32_SNAPSHOT	CTTCGCCAACGCCACCTT	64,6	ja	ja
F_DNA_IGF2BP3_SNAPSHOT	GCTGGAGCAATCTGTAACAGA	58,1	ja	nein
F_IGF2BP3_SNAPSHOT	AGCTTCAATTAAGATTGCTCCAGC	62,3	nein	ja
R_IGF2BP3_SNAPSHOT	CTCACTTTAGCATCTGGTGCTTC	60,4	ja	ja

A.2.5 MID-Primer

Tabelle A.5: Sequenzen der MID-Primer

Primername	Primersequenz (5'-3')
AutMID1	CGTATCGCCTCCCTCGCGCCA TCAGACGAGTGCCT
ButMID1	CTATGCGCCTTGCCAGCCCGC TCAGACGAGTGCCT
AutMID2	CGTATCGCCTCCCTCGCGCCA TCAGACGCTCGACA
ButMID2	CTATGCGCCTTGCCAGCCCGC TCAGACGCTCGACA
AutMID3	CGTATCGCCTCCCTCGCGCCA TCAGAGACGCACTC
ButMID3	CTATGCGCCTTGCCAGCCCGC TCAGAGACGCACTC
AutMID4	CGTATCGCCTCCCTCGCGCCA TCAGAGCACTGTAG
ButMID4	CTATGCGCCTTGCCAGCCCGC TCAGAGCACTGTAG
AutMID5	CGTATCGCCTCCCTCGCGCCA TCAGATCAGACACG
ButMID5	CTATGCGCCTTGCCAGCCCGC TCAGATCAGACACG
AutMID6	CGTATCGCCTCCCTCGCGCCA TCAGATATCGCGAG
ButMID6	CTATGCGCCTTGCCAGCCCGC TCAGATATCGCGAG
AutMID7	CGTATCGCCTCCCTCGCGCCA TCAGCGTGTCTCTA
ButMID7	CTATGCGCCTTGCCAGCCCGC TCAGCGTGTCTCTA
AutMID8	CGTATCGCCTCCCTCGCGCCA TCAGCTCGCGTGTCTC
ButMID8	CTATGCGCCTTGCCAGCCCGC TCAGCTCGCGTGTCTC
AutMID9	CGTATCGCCTCCCTCGCGCCA TCAGTAGTATCAGC
ButMID9	CTATGCGCCTTGCCAGCCCGC TCAGTAGTATCAGC
AutMID10	CGTATCGCCTCCCTCGCGCCA TCAGTCTCTATGCG
ButMID10	CTATGCGCCTTGCCAGCCCGC TCAGTCTCTATGCG
AutMID11	CGTATCGCCTCCCTCGCGCCA TCAGTGATACGCTCT
ButMID11	CTATGCGCCTTGCCAGCCCGC TCAGTGATACGCTCT
AutMID12	CGTATCGCCTCCCTCGCGCCA TCAGTACTGAGCTA
ButMID12	CTATGCGCCTTGCCAGCCCGC TCAGTACTGAGCTA
AutMID13	CGTATCGCCTCCCTCGCGCCA TCAGCATAGTAGTG
ButMID13	CTATGCGCCTTGCCAGCCCGC TCAGCATAGTAGTG
AutMID14	CGTATCGCCTCCCTCGCGCCA TCAGCGAGAGATAC
ButMID14	CTATGCGCCTTGCCAGCCCGC TCAGCGAGAGATAC
AutMID15	CGTATCGCCTCCCTCGCGCCA TCAGATACGACGTA
ButMID15	CTATGCGCCTTGCCAGCCCGC TCAGATACGACGTA
AutMID16	CGTATCGCCTCCCTCGCGCCA TCAGTCACGTACTA
ButMID16	CTATGCGCCTTGCCAGCCCGC TCAGTCACGTACTA
AutMID17	CGTATCGCCTCCCTCGCGCCA TCAGCGTCTAGTAC
ButMID17	CTATGCGCCTTGCCAGCCCGC TCAGCGTCTAGTAC
AutMID18	CGTATCGCCTCCCTCGCGCCA TCAGTCTACGTAGC
ButMID18	CTATGCGCCTTGCCAGCCCGC TCAGTCTACGTAGC
AutMID19	CGTATCGCCTCCCTCGCGCCA TCAGTGTACTACTC
ButMID19	CTATGCGCCTTGCCAGCCCGC TCAGTGTACTACTC
AutMID20	CGTATCGCCTCCCTCGCGCCA TCAGACGACTACAG
ButMID20	CTATGCGCCTTGCCAGCCCGC TCAGACGACTACAG
AutMID21	CGTATCGCCTCCCTCGCGCCA TCAGCGTAGACTAG
ButMID21	CTATGCGCCTTGCCAGCCCGC TCAGCGTAGACTAG
AutMID22	CGTATCGCCTCCCTCGCGCCA TCAGTACGAGTATG
ButMID22	CTATGCGCCTTGCCAGCCCGC TCAGTACGAGTATG
AutMID23	CGTATCGCCTCCCTCGCGCCA TCAGTACTCTCGTG
ButMID23	CTATGCGCCTTGCCAGCCCGC TCAGTACTCTCGTG
AutMID24	CGTATCGCCTCCCTCGCGCCA TCAGTAGAGACGAG
ButMID24	CTATGCGCCTTGCCAGCCCGC TCAGTAGAGACGAG
AutMID25	CGTATCGCCTCCCTCGCGCCA TCAGTCGTCGCTCG
ButMID25	CTATGCGCCTTGCCAGCCCGC TCAGTCGTCGCTCG
AutMID26	CGTATCGCCTCCCTCGCGCCA TCAGACATACGCGT
ButMID26	CTATGCGCCTTGCCAGCCCGC TCAGACATACGCGT
AutMID27	CGTATCGCCTCCCTCGCGCCA TCAGACGCGAGTAT
ButMID27	CTATGCGCCTTGCCAGCCCGC TCAGACGCGAGTAT
AutMID28	CGTATCGCCTCCCTCGCGCCA TCAGACTACTATGT
ButMID28	CTATGCGCCTTGCCAGCCCGC TCAGACTACTATGT

Tabelle A.5: Sequenzen der MID-Primer

Primernam	Primersequenz (5'-3')
AutMID29	CGTATCGCCTCCCTCGCGCCA TCAGACTGTACAGT
ButMID29	CTATGCGCCTTGCCAGCCCGC TCAGACTGTACAGT
AutMID30	CGTATCGCCTCCCTCGCGCCA TCAGAGACTATACT
ButMID30	CTATGCGCCTTGCCAGCCCGC TCAGAGACTATACT
AutMID31	CGTATCGCCTCCCTCGCGCCA TCAGAGCGTCGTCT
ButMID31	CTATGCGCCTTGCCAGCCCGC TCAGAGCGTCGTCT
AutMID32	CGTATCGCCTCCCTCGCGCCA TCAGAGTACGCTAT
ButMID32	CTATGCGCCTTGCCAGCCCGC TCAGAGTACGCTAT
AutMID33	CGTATCGCCTCCCTCGCGCCA TCAGATAGAGTACT
ButMID33	CTATGCGCCTTGCCAGCCCGC TCAGATAGAGTACT
AutMID34	CGTATCGCCTCCCTCGCGCCA TCAGCAGCTACGT
ButMID34	CTATGCGCCTTGCCAGCCCGC TCAGCAGCTACGT
AutMID35	CGTATCGCCTCCCTCGCGCCA TCAGCAGTAGACGT
ButMID35	CTATGCGCCTTGCCAGCCCGC TCAGCAGTAGACGT
AutMID36	CGTATCGCCTCCCTCGCGCCA TCAGCGACGTGACT
ButMID36	CTATGCGCCTTGCCAGCCCGC TCAGCGACGTGACT
AutMID37	CGTATCGCCTCCCTCGCGCCA TCAGTACACGTGAT
ButMID37	CTATGCGCCTTGCCAGCCCGC TCAGTACACGTGAT
AutMID38	CGTATCGCCTCCCTCGCGCCA TCAGTAGTGTAGAT
ButMID38	CTATGCGCCTTGCCAGCCCGC TCAGTAGTGTAGAT
AutMID39	CGTATCGCCTCCCTCGCGCCA TCAGTCGATCACGT
ButMID39	CTATGCGCCTTGCCAGCCCGC TCAGTCGATCACGT
AutMID40	CGTATCGCCTCCCTCGCGCCA TCAGTCTATACTAT
ButMID40	CTATGCGCCTTGCCAGCCCGC TCAGTCTATACTAT
AutMID41	CGTATCGCCTCCCTCGCGCCA TCAGTGACGTATGT
ButMID41	CTATGCGCCTTGCCAGCCCGC TCAGTGACGTATGT
AutMID42	CGTATCGCCTCCCTCGCGCCA TCAGAGTCGAGAGA
ButMID42	CTATGCGCCTTGCCAGCCCGC TCAGAGTCGAGAGA
AutMID43	CGTATCGCCTCCCTCGCGCCA TCAGCGATCGTATA
ButMID43	CTATGCGCCTTGCCAGCCCGC TCAGCGATCGTATA
AutMID44	CGTATCGCCTCCCTCGCGCCA TCAGCGCAGTACGA
ButMID44	CTATGCGCCTTGCCAGCCCGC TCAGCGCAGTACGA
AutMID45	CGTATCGCCTCCCTCGCGCCA TCAGCTACGCTCTA
ButMID45	CTATGCGCCTTGCCAGCCCGC TCAGCTACGCTCTA
AutMID46	CGTATCGCCTCCCTCGCGCCA TCAGTCACGCGAGA
ButMID46	CTATGCGCCTTGCCAGCCCGC TCAGTCACGCGAGA
AutMID47	CGTATCGCCTCCCTCGCGCCA TCAGTGTCACACGA
ButMID47	CTATGCGCCTTGCCAGCCCGC TCAGTGTCACACGA

A.3 Ergebnisse der Genotypisierung

A.3.1 Genotypisierung der 22 Monozyten

Tabelle A.6: Genotypisierung von 14414_1_hg19(*ASRGL1*), 15224_1_hg19 (*PARP11*) und 19100_1_hg19 (*PDXDC1*) in 22 Monozyten-Proben

Probe	<i>ASRGL1</i> rs11231058	<i>PARP11</i> rs12319851	<i>PDXDC1</i> rs9928601
R1	*	A/G	A/C
R2	A/G	A/G	A/C
R3	A/A	A/G	A/C
R4	A/G	A/G	A/A
R5	*	A/A	C/C
R6	A/G	A/A	A/C
R7	G/G	A/G	A/C
R8	A/G	A/A	A/C
R9	A/A	A/G	A/C
R10	A/G	A/G	A/C
R11	A/G	A/G	A/C
R12	*	G/G	*
R13	G/G	A/G	A/C
R14	A/G	G/G	A/C
R15	A/G	A/A	C/C
R16	G/G	A/G	C/C
R17	A/G	A/A	A/C
P1	A/G	A/A	A/A
P2	A/G	A/A	A/A
P3	A/G	A/A	A/A
K1	A/G	A/G	A/C
K2	A/G	A/G	A/C

*PCR nicht funktioniert

Tabelle A.7: Genotypisierung von 2391_1_hg19, 6141_1_hg19, 9224_1_hg19, 9261_1_hg19, 9377_1_hg19 und 9473_1_hg19 in 22 Monozyten-Proben

Probe	2391_1_hg19 rs11806449	6141_1_hg19 rs1849135	9224_1_hg19 rs6959065	9261_1_hg19 rs10228640	9377_1_hg19 rs140384299	9473_1_hg19 rs6243194
R1	G/G	A/A	G/G	A/A	T/T	A/A
R2	G/G	A/A	G/G	A/A	T/T	A/A
R3	G/A	A/C	A/G	A/A	T/T	A/A
R4	G/A	C/C	G/G	A/A	T/T	A/A
R5	G/G	C/C	A/G	A/G	*	A/A
R6	G/G	A/C	A/G	A/A	*	A/A
R7	G/G	A/C	A/G	A/A	T/T	A/A
R8	G/A	C/C	A/G	A/A	T/T	A/G
R9	G/A	C/C	G/G	A/G	T/T	A/A
R10	G/G	C/C	A/G	A/A	T/T	A/G
R11	G/A	A/C	G/G	A/A	T/T	A/A
R12	G/G	C/C	G/G	A/A	T/T	A/A
R13	G/G	A/C	G/G	A/G	T/T	A/A
R14	A/A	A/C	G/G	A/A	T/T	A/A
R15	G/A	C/C	A/A	A/A	T/T	A/A
R16	A/A	A/C	G/G	A/A	T/T	A/A
R17	n.s.	A/A	n.s.	n.s.	T/T	A/A
P1	n.s.	n.s.	n.s.	n.s.	T/T	A/A
P2	n.s.	n.s.	n.s.	n.s.	T/T	A/A
P3	n.s.	n.s.	n.s.	n.s.	T/T	n.s.
K1	n.s.	n.s.	n.s.	n.s.	T/T	n.s.
K2	n.s.	n.s.	n.s.	n.s.	T/T	n.s.

*PCR nicht funktioniert; n.s. nicht sequenziert

Tabelle A.8: Genotypisierung von 13250_1_hg19, 15400_1_hg19, 15512_1_hg19, 16448_1_hg19 und 16458_1_hg19 in 22 Monozyten-Proben

Probe	13250_1_hg19 rs61869800	15400_1_hg19 rs670519	15512_1_hg19 rs77023346	16448_1_hg19 rs4770130	16458_1_hg19 rs7333053
R1	G/G	A/A	T/T	G/G	C/C
R2	G/G	A/A	T/T	G/G	C/C
R3	G/A	A/A	T/T	G/A	C/C
R4	G/G	A/G	T/T	G/G	C/C
R5	G/A	A/G	T/T	G/A	C/C
R6	G/G	*	T/T	G/G	*
R7	G/G	A/A	T/T	G/G	C/C
R8	G/G	*	T/T	G/A	C/C
R9	G/A	A/A	T/T	A/A	C/C
R10	G/G	A/A	T/T	G/G	C/C
R11	G/G	*	T/T	G/A	C/C
R12	G/G	A/G	T/T	G/A	C/C
R13	G/A	A/A	T/T	G/G	C/C
R14	G/G	A/G	T/T	G/G	C/C
R15	G/G	A/A	*	G/G	C/C
R16	G/A	A/A	T/T	G/A	C/C
R17	G/G	A/G	T/T	n.s.	C/C
P1	G/G	A/A	T/T	n.s.	C/C
P2	G/G	A/A	T/T	n.s.	C/T
P3	n.s.	A/A	T/T	n.s.	n.s.
K1	n.s.	A/A	T/T	n.s.	n.s.
K2	n.s.	A/A	T/T	n.s.	n.s.

*PCR nicht funktioniert; n.s. nicht sequenziert

Tabelle A.9: Genotypisierung von 17031_1_hg19, 20403_1_hg19, 23548_1_hg19, 24982_1_hg19 und 26859_1_hg19 in 22 Monozyten-Proben

Probe	17031_1_hg19 rs7141760	20403_1_hg19 rs11078314	23548_1_hg19 rs12975447	24982_1_hg19 45612231	26859_1_hg19 kein SNP
R1	C/C	C/G	G/G	T/T	—
R2	G/G	C/C	T/G	T/T	—
R3	C/C	C/G	G/G	T/T	—
R4	C/G	G/G	G/G	T/T	—
R5	C/C	C/G	G/G	T/T	—
R6	C/C	*	G/G	T/T	—
R7	C/C	G/G	G/G	T/T	—
R8	C/C	C/G	G/G	T/T	—
R9	*	C/G	G/G	T/T	—
R10	C/C	G/G	G/G	T/T	—
R11	C/C	G/G	G/G	T/T	—
R12	C/C	C/G	G/G	T/T	—
R13	C/C	G/G	G/G	T/T	—
R14	C/G	G/G	G/G	T/T	—
R15	C/C	C/C	T/G	*	—
R16	C/C	C/G	G/G	T/T	—
R17	C/C	C/G	G/G	T/T	—
P1	n.s.	C/C	n.s.	T/T	—
P2	n.s.	C/C	n.s.	T/T	—
P3	n.s.	n.s.	n.s.	T/T	—
K1	n.s.	n.s.	n.s.	T/T	—
K2	n.s.	n.s.	n.s.	T/T	—

*PCR nicht funktioniert; n.s. nicht sequenziert

A.4 Ausführliche Ergebnisse der Methylierungsanalysen

A.4.1 Methylierungsgrad von 104 human/nicht-murinen intronischen CGIs mit zusätzlichen Treffern im Genom (Methylom 1 und 2)

Tabelle A.10: Methylierungsdaten von 104 human/nicht-murinen intronischen CGIs mit zusätzlichen Treffern im Genom (Methylom 1 und 2). Diese Tabelle wurde schon in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

CGI ID	Methylom 1		Methylom 2	
	Meth.(%)	Abd. (#)	Meth.(%)	Abd.(#)
334_1_hg19	9	6	10	5
1094_1_hg19	94	12	84	6
1147_1_hg19	99	26	97	27
1227_1_hg19	99	14	98	5
1255_1_hg19	86	13	97	10
1287_1_hg19	97	17	98	9
1330_1_hg19	0	6	0	4
1362_1_hg19	99	10	98	5
1720_1_hg19	1	15	1	6
1909_1_hg19	0	14	4	7
1911_1_hg19	81	18	58	6
2392_1_hg19	97	14	94	5
2708_1_hg19	97	22	97	5
3524_1_hg19	0	3	0	4
3783_1_hg19	97	12	99	8
3926_1_hg19	96	23	97	12
4213_1_hg19	96	8	99	8
4675_1_hg19	73	19	79	15
4754_1_hg19	32	9	17	6
5128_1_hg19	3	4	2	5
5148_1_hg19	90	11	97	6
5707_1_hg19	96	12	96	10
5734_1_hg19	98	16	99	7
5735_1_hg19	97	16	98	9
5774_1_hg19	95	17	98	10
5895_1_hg19	97	23	98	15
6119_1_hg19	98	30	97	27
6360_1_hg19	0	8	3	3
6681_1_hg19	92	22	94	6
6781_1_hg19	93	17	97	5
6782_1_hg19	97	25	95	13
6931_1_hg19	98	22	97	24
6952_1_hg19	97	13	96	6
7353_1_hg19	99	15	90	4
7430_1_hg19	97	26	96	16
7480_1_hg19	83	15	97	7
8436_1_hg19	97	14	97	9
9007_1_hg19	94	12	96	6
9008_1_hg19	97	20	97	9
9009_1_hg19	78	8	77	2
9933_1_hg19	98	16	98	11
9934_1_hg19	95	15	93	8
10233_1_hg19	96	9	96	7
10850_1_hg19	97	16	98	13
11376_1_hg19	9	2	1	2
11477_1_hg19	100	9	99	5
11479_1_hg19	96	15	98	7
11481_1_hg19	97	12	93	6
11497_1_hg19	98	19	96	10
11511_1_hg19	96	23	97	18
11525_1_hg19	99	14	95	12
12025_1_hg19	95	11	84	5
13648_1_hg19	89	7	88	4
14414_1_hg19	55	10	49	9
14676_1_hg19	3	6	0	3
14830_1_hg19	98	20	99	13
14831_1_hg19	98	23	99	10
15161_1_hg19	2	7	1	5
15180_1_hg19	3	5	2	4
15205_1_hg19	84	12	76	6
15224_1_hg19	42	11	46	8
15290_1_hg19	23	14	20	11
15352_1_hg19	98	17	97	8
15356_1_hg19	96	17	96	14
15798_1_hg19	96	12	92	6
16205_1_hg19	97	10	96	5
16226_1_hg19	98	12	96	9
16253_1_hg19	95	21	96	6
16256_1_hg19	87	21	90	7
16633_1_hg19	97	15	97	7

Tabelle A.10: Methylierungsdaten von 104 human/nicht-murinen intronischen CGIs mit zusätzlichen Treffern im Genom (Methylom 1 und 2). Diese Tabelle wurde schon in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

CGI ID	Methylom 1		Methylom 2	
	Meth.(%)	Abd. (#)	Meth.(%)	Abd.(#)
16634_1_hg19	63	13	72	8
16694_1_hg19	97	23	96	18
17304_1_hg19	96	15	96	10
17484_1_hg19	99	13	96	6
17523_1_hg19	98	12	97	6
17547_1_hg19	1	16	1	9
18132_1_hg19	98	17	93	9
18136_1_hg19	95	23	97	12
18308_1_hg19	95	9	98	5
18512_1_hg19	96	14	95	7
19100_1_hg19	75	11	79	6
19854_1_hg19	96	22	96	16
19870_1_hg19	58	15	59	8
19877_1_hg19	98	14	96	4
19878_1_hg19	91	12	88	10
19881_1_hg19	96	19	95	8
19883_1_hg19	97	7	93	5
20085_1_hg19	96	11	92	5
20632_1_hg19	22	7	2	2
20636_1_hg19	58	6	74	5
20657_1_hg19	98	15	97	10
21034_1_hg19	95	14	94	8
21352_1_hg19	94	20	92	11
21920_1_hg19	2	16	1	5
22285_1_hg19	91	9	88	5
25767_1_hg19	70	17	75	13
26094_1_hg19	96	15	94	14
26614_1_hg19	96	17	97	6
26734_1_hg19	0	8	0	4
26772_1_hg19	95	10	86	5
26785_1_hg19	98	6	100	2
26912_1_hg19	97	9	98	4
26939_1_hg19	3	5	0	1
27070_1_hg19	97	12	96	16

A.4.2 Histogramme der Methylierung der einzelnen Reads der 14 Kandidaten-CGIs (Methylom 1 und 2)

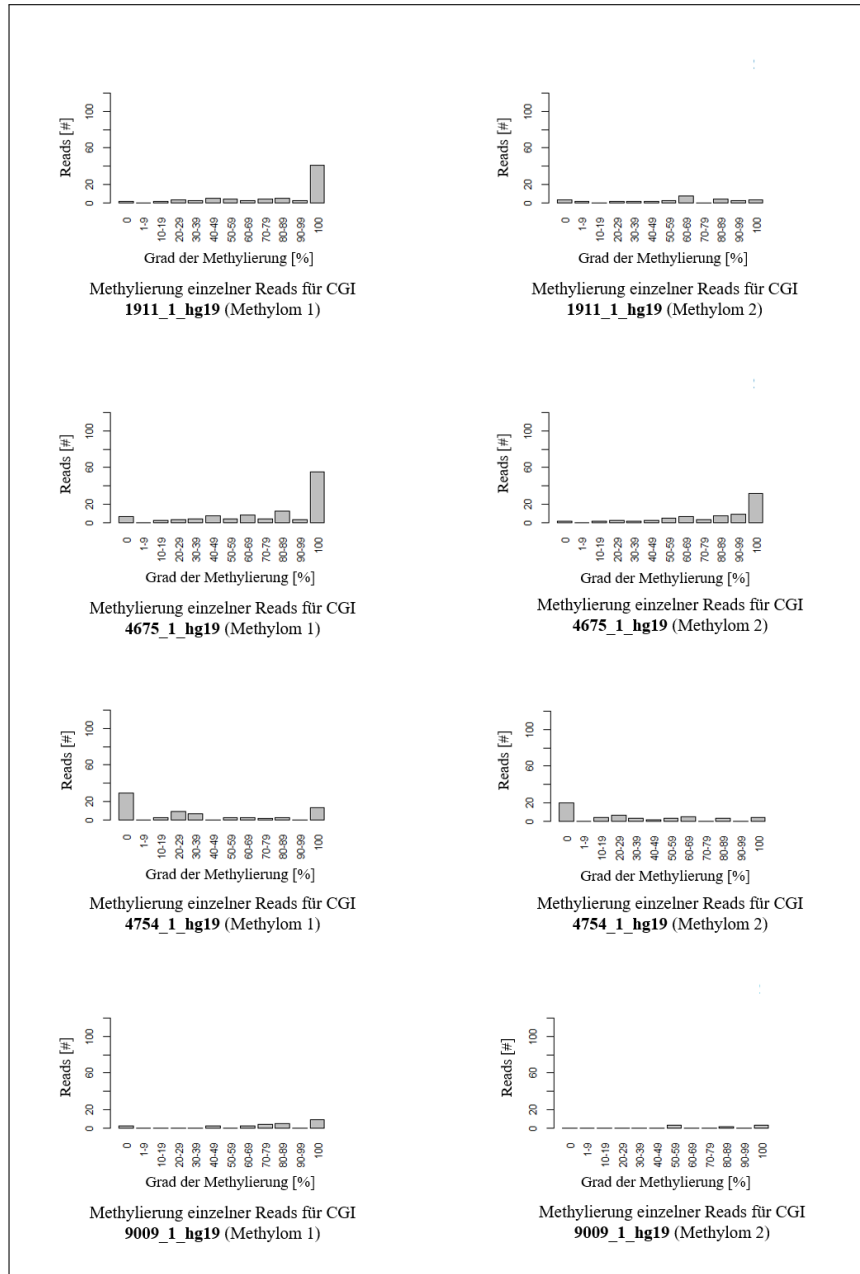


Abbildung A.1: Histogramme über die Methylierung einzelner Reads für human/nicht-murin intronische CGIs mit einem mittleren Methylierungsniveau - Part 1

Für jedes CGI sind zwei Histogramme angegeben, die die Verteilung von Reads (y-Achse) in Abhängigkeit von ihrer Methylierung (x-Achse) für Methylom 1 und Methylom 2 zeigen. Diese Abbildungen wurden bereits in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

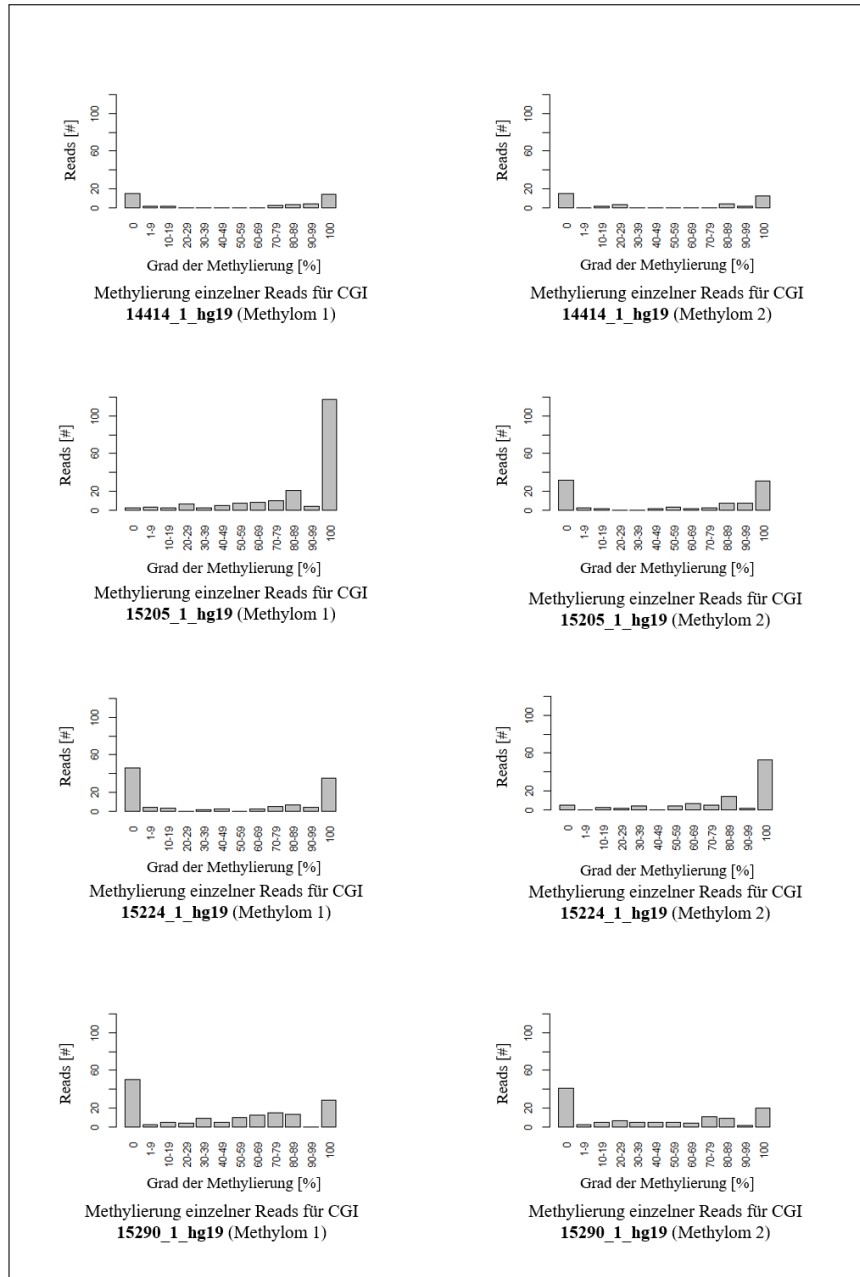


Abbildung A.2: Histogramme über die Methylierung einzelner Reads für human/nicht-murin intronische CGIs mit einem mittleren Methylierungsniveau - Part 2

Für jedes CGI sind zwei Histogramme angegeben, die die Verteilung von Reads (y-Achse) in Abhängigkeit von ihrer Methylierung (x-Achse) für Methylom 1 und Methylom 2 zeigen. Diese Abbildungen wurden bereits in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

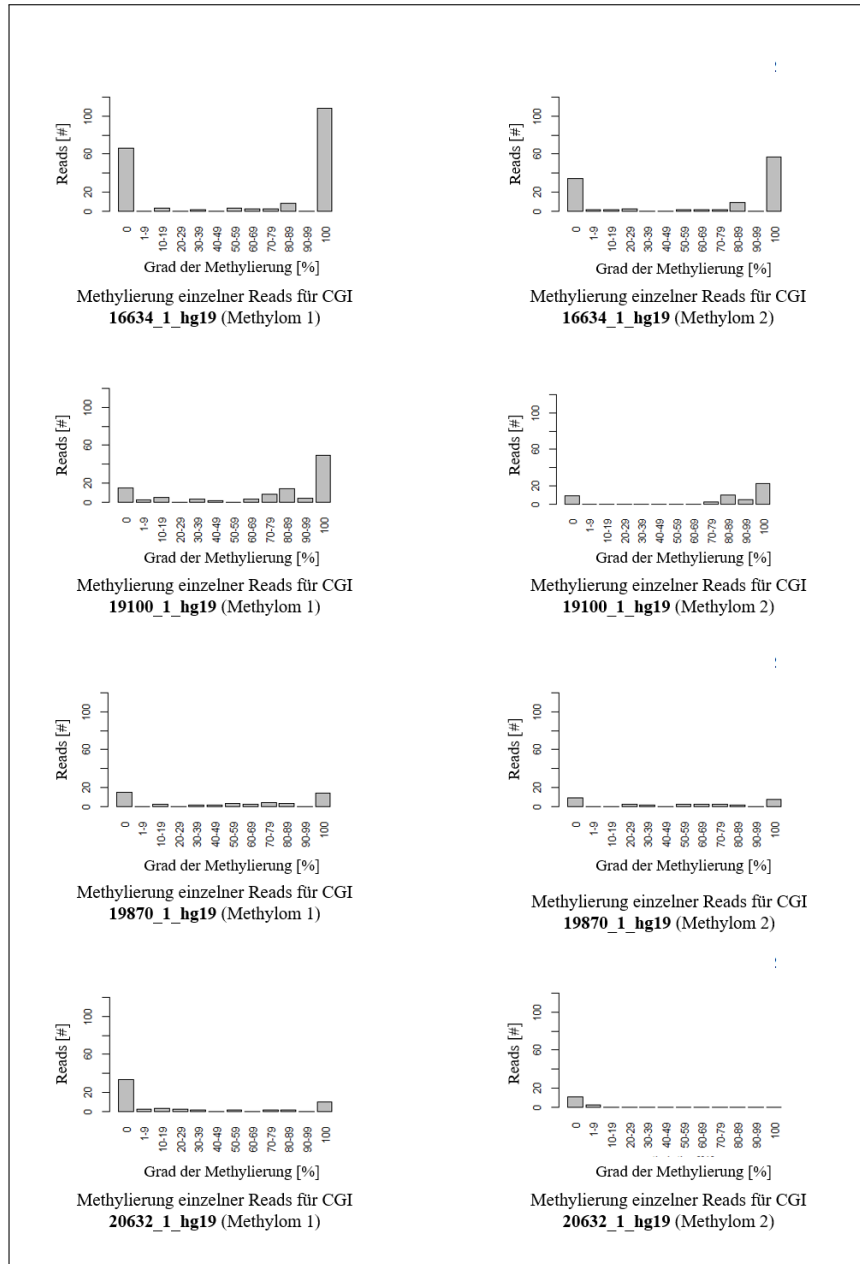


Abbildung A.3: Histogramme über die Methylierung einzelner Reads für human/nicht-murin intronische CGIs mit einem mittleren Methylierungsniveau - Part 3

Für jedes CGI sind zwei Histogramme angegeben, die die Verteilung von Reads (y-Achse) in Abhängigkeit von ihrer Methylierung (x-Achse) für Methylom 1 und Methylom 2 zeigen. Diese Abbildungen wurden bereits in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

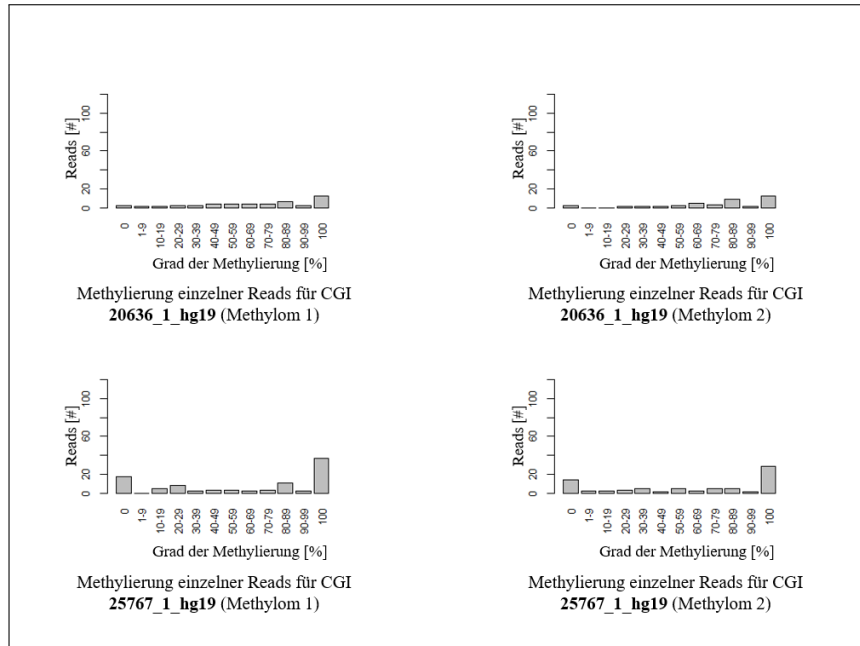


Abbildung A.4: Histogramme über die Methylierung einzelner Reads für human/nicht-murin intronische CGIs mit einem mittleren Methylierungsniveau - Part 4

Für jedes CGI sind zwei Histogramme angegeben, die die Verteilung von Reads (y-Achse) in Abhängigkeit von ihrer Methylierung (x-Achse) für Methylom 1 und Methylom 2 zeigen. Diese Abbildungen wurden bereits in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

A.4.3 Methylierungsgrad der CGIs 14414_1_hg19 (*ASRGL1*), 15224_1_hg19 (*PARP11*) und 19100_1_hg19 (*PDXDC1*) in Monozyten

Tabelle A.11: Methylierungsgrad des CGIs 14414_1_hg19(*ASRGL1*) in Monozyten

Probe	Ø Meth. (%)	CpGs (#)	Reads (#)	Allel a			Allel b		
				Allel	Ø Meth. (%)	Reads (#)	Allel	Ø Meth. (%)	Reads (#)
R2	89,5%	34	518	A	78,5%	99	G	92,9%	386
R4	93,0%	34	796	A	94,0%	414	G	92,8%	261
R6	96,1%	34	389	A	96,9%	181	G	95,8%	196
R8	89,6%	34	696	A	86,4%	290	G	92,0%	385
R10	88,2%	34	239	A	91,0%	151	G	85,1%	64
R11	78,4%	34	1282	A	82,1%	577	G	77,1%	598
R14	94,6%	34	1096	A	96,7%	519	G	93,9%	418
R15	77,7%	34	2042	A	92,5%	1254	G	35,9%	407
R17	92,6%	34	1537	A	93,5%	742	G	93,0%	553
P1	75,9%	34	5483	A	78,4%	3197	G	72,8%	1853
P2	72,5%	34	4304	A	89,7%	1904	G	54,4%	1811
P3	89,7%	34	4798	A	93,0%	3126	G	82,1%	1209
K1	74,9%	34	943	A	91,7%	380	G	62,7%	491
K2	89,6%	34	3847	A	90,9%	1707	G	89,4%	1752

Tabelle A.12: Methylierungsgrad des CGIs 15224_1_hg19 (*PARP11*) in Monozyten

Probe	Ø Meth. (%)	CpGs (#)	Reads (#)	Allel a			Allel b		
				Allel	Ø Meth. (%)	Reads (#)	Allel	Ø Meth. (%)	Reads (#)
R1	83,1%	30	1270	A	82,5%	496	G	83,8%	734
R2	67,5%	30	705	A	59,0%	339	G	75,1%	356
R3	78,8%	30	1357	A	76,8%	689	G	80,8%	628
R4	79,7%	30	917	A	77,6%	372	G	81,1%	501
R7	71,7%	30	1151	A	72,7%	541	G	70,6%	566
R9	69,9%	30	1751	A	64,8%	796	G	74,0%	873
R10	84,4%	30	1467	A	86,5%	647	G	82,4%	721
R11	73,9%	30	978	A	75,3%	441	G	72,5%	507
R13	75,6%	30	1340	A	79,0%	673	G	71,4%	605
R16	79,7%	30	2066	A	78,9%	943	G	79,9%	969
K1	72,5%	30	1121	A	72,3%	558	G	72,6%	550
K2	72,3%	30	9522	A	771,5%	4760	G	72,7%	4348

Tabelle A.13: Methylierungsgrad des CGIs 19100_1_hg19(*PDXDC1*) in Monozyten

Probe	Ø Meth. (%)	CpGs (#)	Reads (#)	Allel a			Allel b		
				Allel	Ø Meth. (%)	Reads (#)	Allel	Ø Meth. (%)	Reads (#)
R1	67,6%	36	520	A	90,7%	380	C	4,0%	138
R2	61,0%	36	471	A	65,7%	221	C	56,6%	246
R3	70,2%	36	344	A	86,6%	168	C	54,5%	176
R6	60,6%	36	511	A	75,4%	385	C	14,6%	123
R7	54,0%	36	456	A	58,2%	117	C	52,7%	338
R8	73,3%	36	380	A	77,1%	285	C	61,5%	94
R9	55,0%	36	740	A	89,9%	302	C	30,5%	435
R10	79,7%	36	548	A	90,4%	250	C	70,5%	292
R11	68,9%	36	555	A	79,0%	373	C	48,2%	182
R13	61,4%	36	618	A	89,3%	254	C	42,1%	362
R14	76,6%	36	343	A	70,2%	169	C	83,2%	172
R17	76,2%	36	752	A	89,7%	371	C	62,9%	379
K1	58,7%	36	1784	A	86,5%	935	C	27,9%	847
K2	52,6%	36	1492	A	89,6%	705	C	18,6%	778

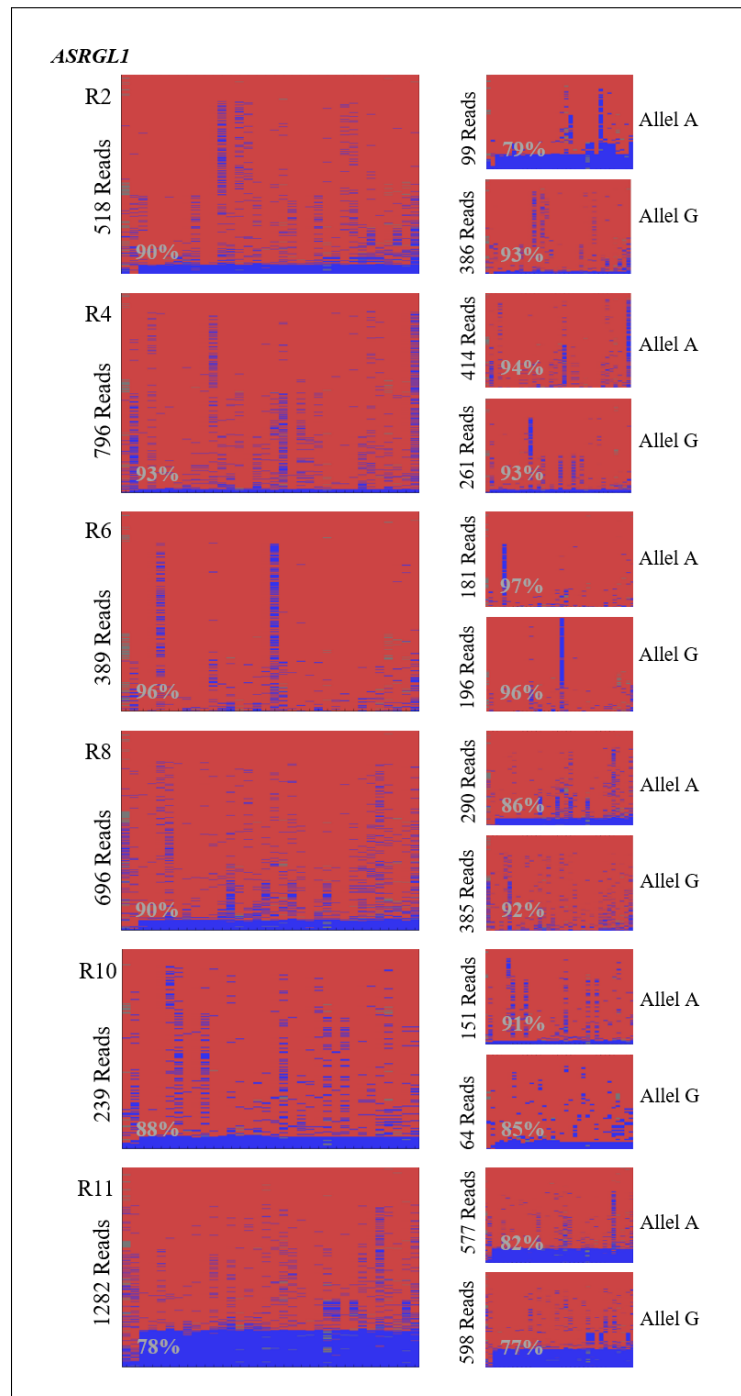


Abbildung A.5: Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozygoten Individuen - Part 1

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.1.2.2 Abbildung 3.2. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 1 zeigt die Methylierungsmuster von CGI 14414_1_hg19 (*ASRGL1*), wobei die ersten zwei CpGs nicht zum CGI gehören.

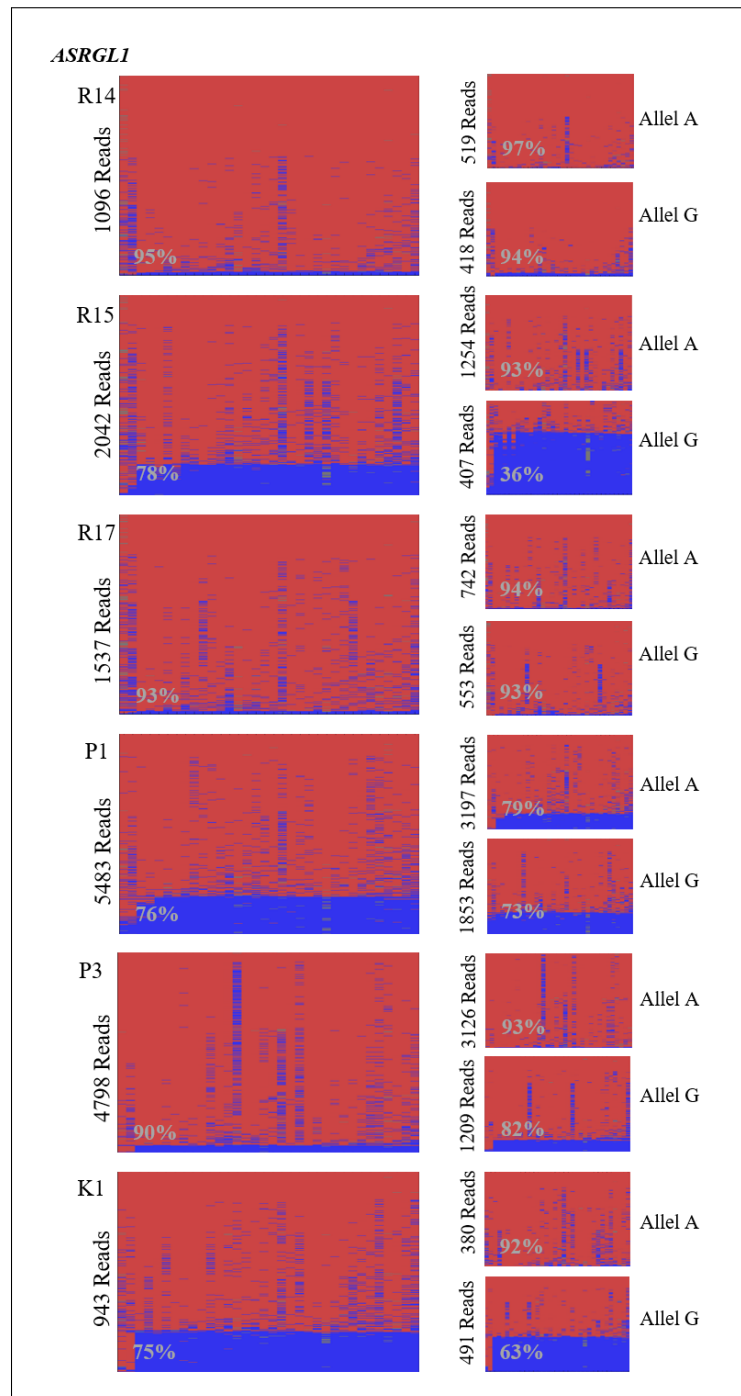


Abbildung A.6: Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozygoten Individuen - Part 2

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.1.2.2 Abbildung 3.2. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 2 zeigt die Methylierungsmuster von CGI 14414_1_hg19 (*ASRGL1*), wobei die ersten zwei CpGs nicht zum CGI gehören.

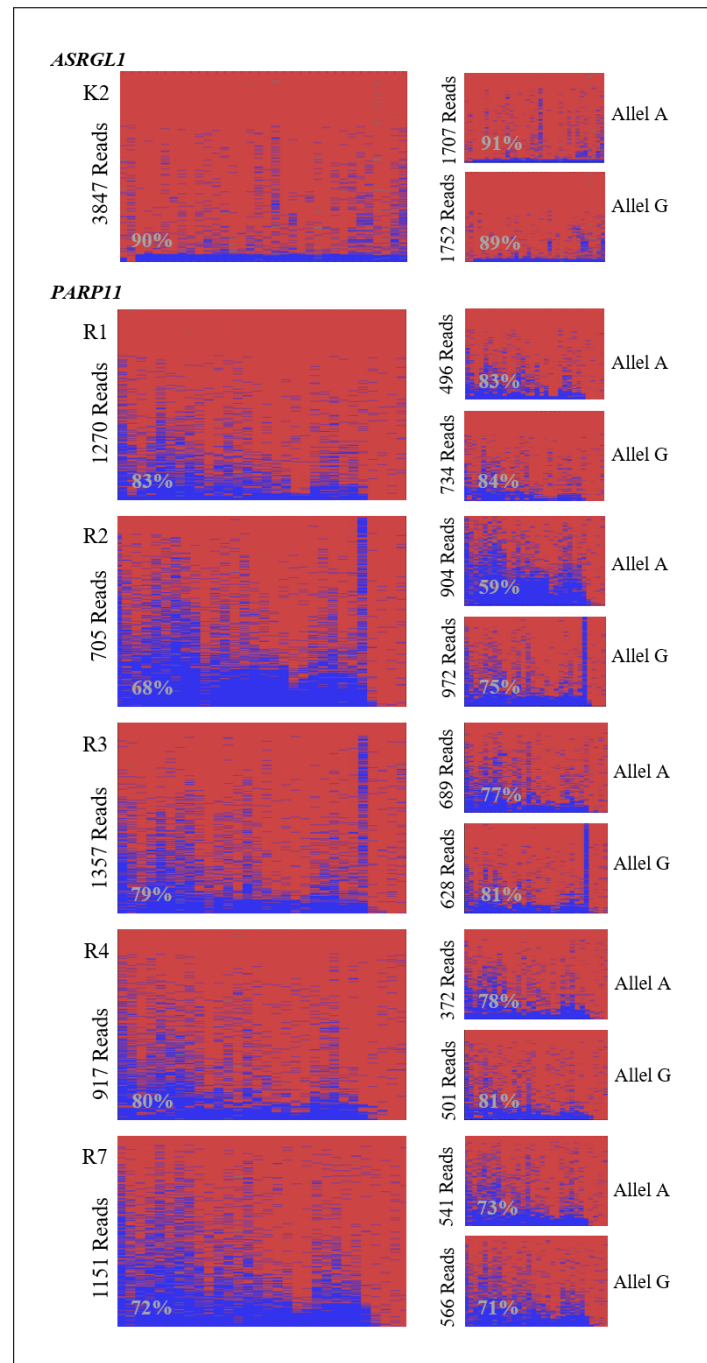


Abbildung A.7: Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozygoten Individuen - Part 3

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.1.2.2 Abbildung 3.2. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 3 zeigt ein Methylierungsmuster von CGI 14414_1_hg19 (*ASRGL1*), wobei die ersten zwei CpGs nicht zum CGI gehören und Methylierungsplots von CGI 15224_1_hg19 (*PARP11*).

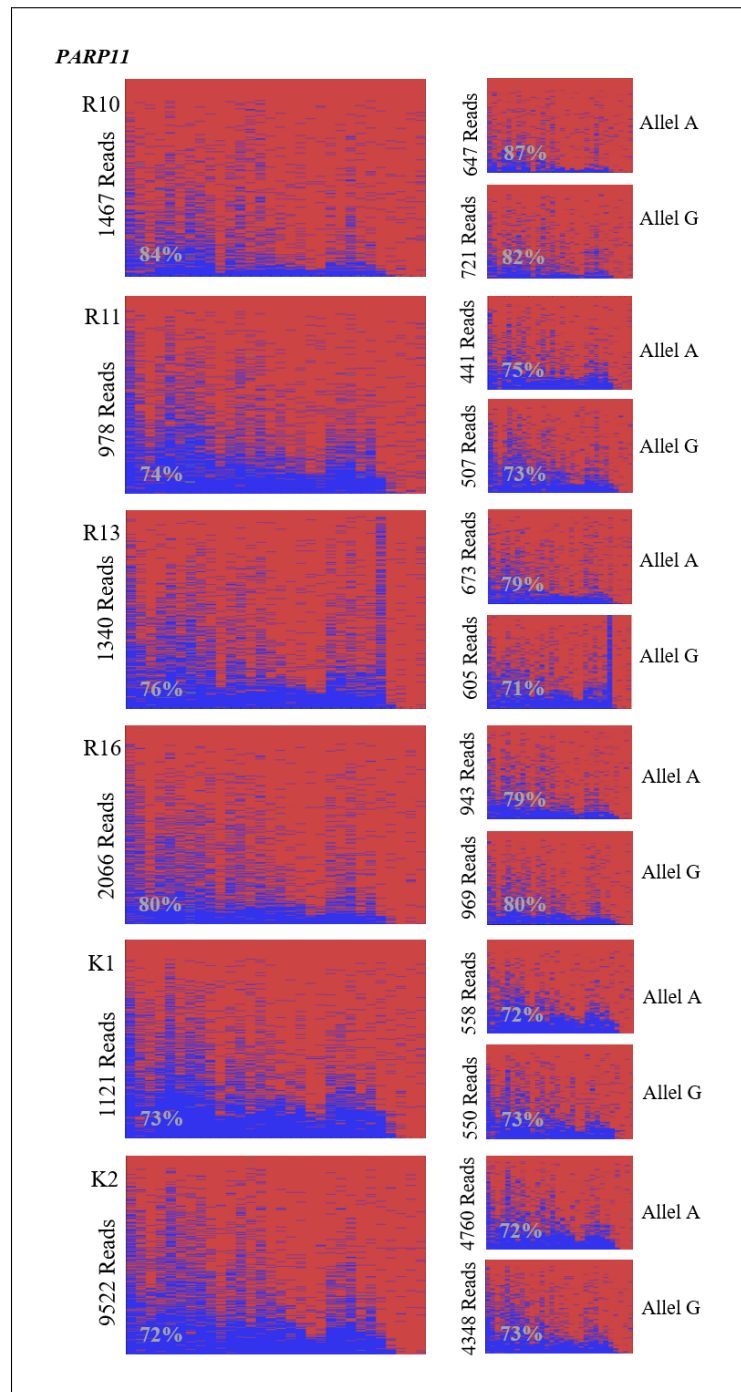


Abbildung A.8: Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozygoten Individuen - Part 4

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.1.2.2 Abbildung 3.2. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 4 zeigt die Methylierungsmuster von CGI 15224_1_hg19 (*PARP11*).

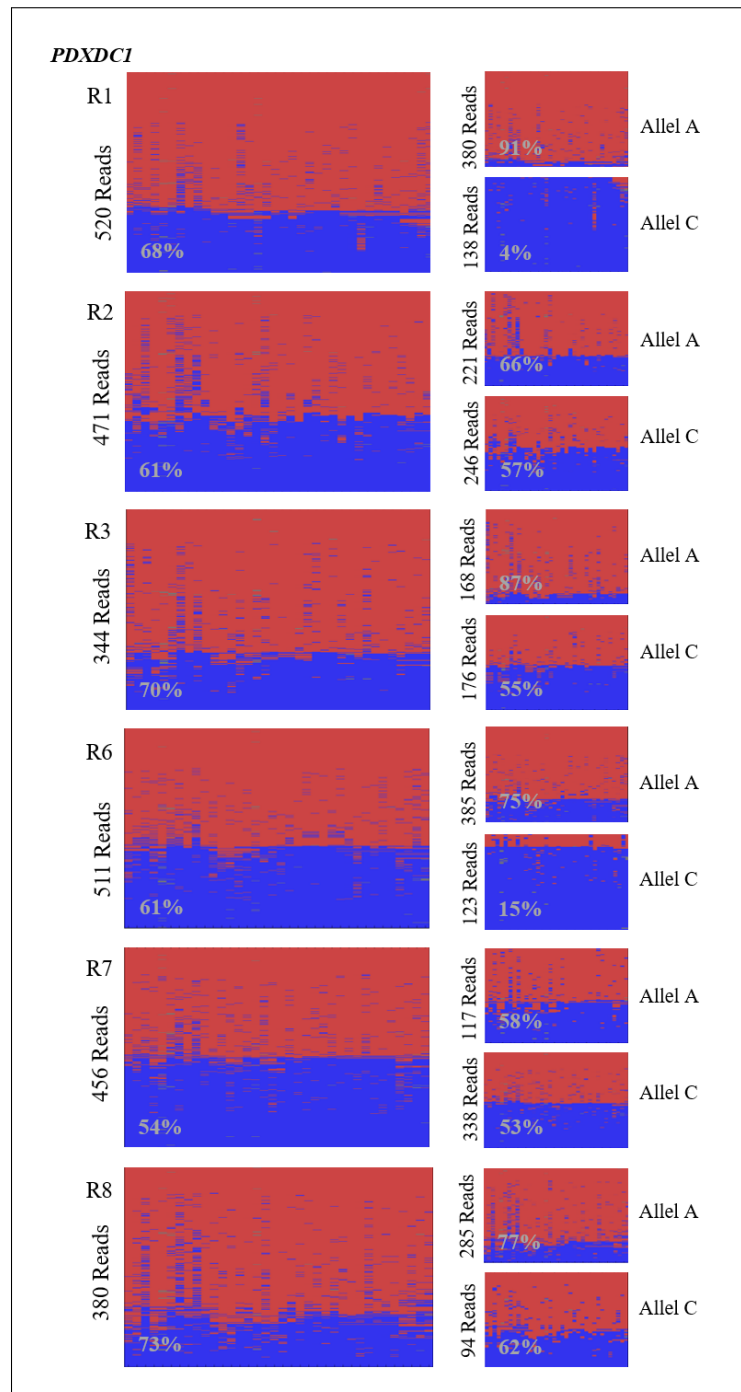


Abbildung A.9: Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozygoten Individuen - Part 5

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.1.2.2 Abbildung 3.2. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 5 zeigt die Methylierungsmuster von CGI 19100_1_hg19 (*PDXDC1*), wobei die letzten vier CpGs nicht zum CGI gehören.

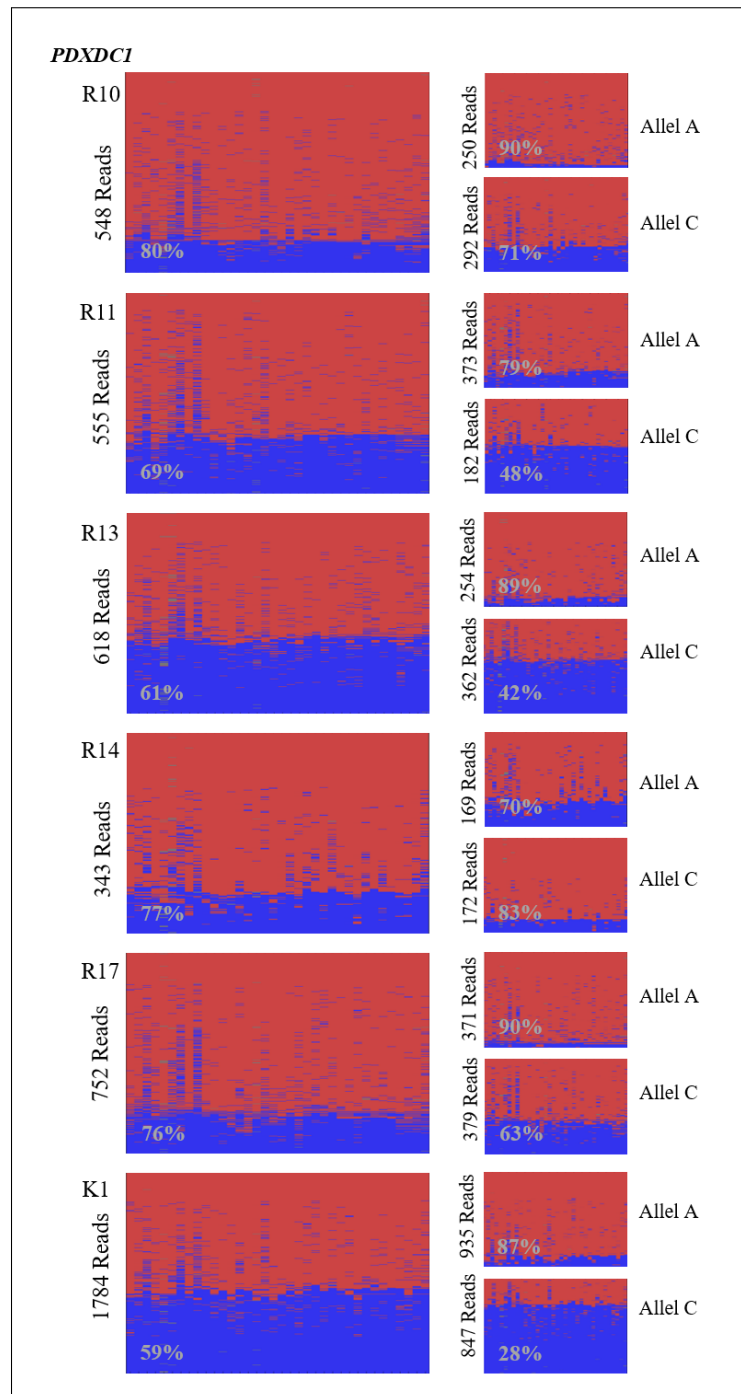


Abbildung A.10: Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozygoten Individuen - Part 6

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.1.2.2 Abbildung 3.2. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 6 zeigt die Methylierungsmuster von CGI 19100_1_hg19 (*PDXDC1*), wobei die letzten vier CpGs nicht zum CGI gehören.

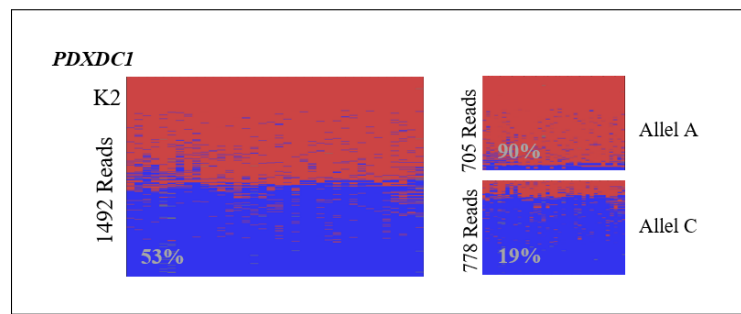


Abbildung A.11: Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Monozyten von heterozygoten Individuen - Part 7

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.1.2.2 Abbildung 3.2. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 7 zeigt die Methylierungsmuster von CGI 19100_1_hg19 (*PDXDC1*), wobei die letzten vier CpGs nicht zum CGI gehören.

A.4.4 Methylierungsgrad der CGIs 14414_1_hg19 (*ASRGL1*), 15224_1_hg19 (*PARP11*) und 19100_1_hg19 (*PDXDC1*) in Blut

Tabelle A.14: Methylierungsgrad der CGIs 14414_1_hg19 (*ASRGL1*), 15224_1_hg19 (*PARP11*) und 19100_1_hg19 (*PDXDC1*) in Blut

Gen	Ø Meth. (%)	CpGs (#)	Reads (#)	Allel a			Allel b		
				Allel	Ø Meth. (%)	Reads (#)	Allel	Ø Meth. (%)	Reads (#)
<i>ASRGL1</i>	87,7	34	4542	A	88,7	2703	G	86,7	1602
<i>ASRGL1</i>	84,0	34	1326	A	91,7	533	G	78,2	721
<i>PARP11</i>	77,2	30	2045	A	76,5	940	G	77,3	994
<i>PARP11</i>	71,1	30	1943	A	72,7	904	G	69,1	972
<i>PARP11</i>	70,8	30	2912	A	67,9	1482	G	73,5	1350
<i>PDXDC1</i>	51,1	36	2339	A	80,6	1217	C	19,1	1118
<i>PDXDC1</i>	59,5	36	2714	A	67,7	1902	C	40,2	798
<i>PDXDC1</i>	55,2	36	1923	A	59,9	1012	C	53,5	884
<i>PDXDC1</i>	51,6	36	1853	A	51,9	953	C	51,7	887

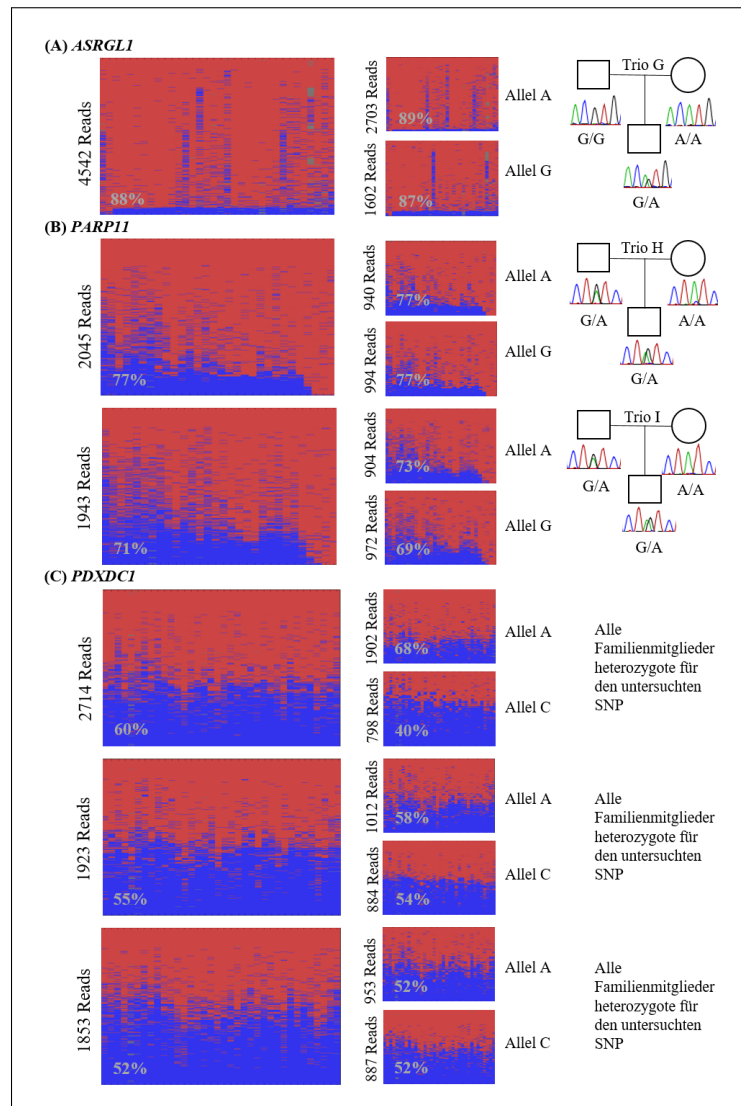


Abbildung A.12: Methylierungsanalyse einzelner CpGs der drei intronischen CGIs (14414_1_hg19, 15224_1_hg19 und 19100_1_hg19) in Blut von heterozygoten Individuen, die teilweise informativ in Bezug auf die elterliche Herkunft sind

Die Abbildung zeigt die zusätzlich untersuchten Proben von Abschnitt 3.1.2.2 Abbildung 3.3. Die Amplikons decken jeweils nur einen Teil der CGIs ab bzw. können noch flankierende CpGs einschließen. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Rechts neben den Plots ist der Stammbaum der jeweiligen Familie angegeben, der die Elektropherogramme des analysierten SNPs für die Familienmitglieder enthält. **(A)** Methylierungsmuster von CGI 14414_1_hg19 (*ASRGL1*), wobei die ersten zwei CpGs nicht zum CGI gehören. **(B)** Methylierungsmuster von CGI 15224_1_hg19 (*PARP11*). **(C)** Methylierungsmuster von CGI 19100_1_hg19 (*PDXDC1*), wobei die letzten vier CpGs nicht zum CGI gehören. Ein Stammbaum ist hier nicht angegeben, da alle Familienmitglieder heterozygot für den untersuchten SNP sind.

A.4.5 Charakteristik von R-CGIs anhand dem Expressionslevel, der genomische Position und dem Grad der Methylierung

Tabelle A.15: Expressionslevel aller R-CGIs in Abhängigkeit ihrer Methylierung

	Expressionslevel							
	exprimiert		exprimiert kein ORF		Pseudogen		gesamte Anzahl	
Anzahl von Retrokopien	934		188		197		1319	
Methylierung der CGIs	(#)	(%)	(#)	(%)	(#)	(%)	(#)	(%)
CGIs m < 20%	273	29	84	45	27	14	384	29
CGIs 20% \geq m \leq 80%	125	13	30	16	40	20	195	15
CGIs m > 80%	436	47	45	24	107	54	588	45
keine Abdeckung	100	11	29	15	23	12	152	12

Die Tabelle zeigt den Methylierungsgrad der R-CGIs in Abhängigkeit der Expression der Retrokopie (UCSC Browser).

Tabelle A.16: Genomische Lage (RefSeq) aller R-CGIs in Abhängigkeit ihrer Methylierung

	Genomische Lage					
	intergenic		intragenic		gesamte Anzahl	
Anzahl von Retrokopien	660		659		1319	
Methylierung der CGIs	(#)	(%)	(#)	(%)	(#)	(%)
CGIs m < 20%	119	18	265	40	384	29
CGIs 20% \geq m \leq 80%	124	19	68	10	192	15
CGIs m > 80%	313	47	278	42	591	45
keine Abdeckung	104	16	48	7	152	12

Die Tabelle zeigt den Methylierungsgrad der R-CGIs in Abhängigkeit ihrer genomischen Lage (RefSeq).

Tabelle A.17: Genomische Lage (RefSeq) aller R-CGIs in Abhängigkeit ihres Expressionslevels

	Genomische Lage					
	intergenic		intragenic		gesamte Anzahl	
Anzahl von Retrokopien	660		659		1319	
Art der Expression	(#)	(%)	(#)	(%)	(#)	(%)
alle vier exprimierte Arten	395	60	539	82	934	71
exprimiert	8	1	10	2	18	1
schwach exprimiert	327	50	109	17	436	33
start exprimiert	42	6	119	18	161	12
in exprimierten Bereich inseriert	18	3	301	46	319	24
exprimiert kein ORF	90	14	98	15	188	14
Pseudogen	175	27	22	3	197	15

Die Tabelle zeigt die genomische Lage (RefSeq) der R-CGIs in Abhängigkeit ihrer Expression (UCSC Browser).

Tabelle A.18: Expressionslevel aller *intergenic* R-CGIs in Abhängigkeit ihrer Methylierung

	Expressionslevel							
	exprimiert		in exprimierten Bereich inseriert		stark exprimiert		stark exprimiert kein ORF	
Anzahl von Retrokopien	8		18		42		14	
Methylierung der CGIs	(#)	(%)	(#)	(%)	(#)	(%)	(#)	(%)
CGIs m < 20%	2	25	3	17	9	21	7	50
CGIs 20% ≥ m ≤ 80%	0	0	2	11	9	21	1	7
CGIs m > 80%	4	50	8	44	21	50	1	7
keine Abdeckung	2	25	5	28	3	7	5	36

	Expressionslevel							
	schwach exprimiert		schwach exprimiert kein ORF		Pseudogen		gesamte Anzahl	
Anzahl von Retrokopien	327		76		175		660	
Methylierung der CGIs	(#)	(%)	(#)	(%)	(#)	(%)	(#)	(%)
CGIs m < 20%	49	15	26	34	22	13	118	18
CGIs 20% ≥ m ≤ 80%	65	20	11	14	38	22	126	19
CGIs m > 80%	162	50	23	30	93	53	312	47
keine Abdeckung	51	16	16	21	22	13	104	16

Die Tabelle zeigt den Methylierungsgrad aller *intergenic* R-CGIs in Abhängigkeit der Expression der Retrokopie (UCSC Browser).

Tabelle A.19: Expressionslevel aller *intragenic* R-CGIs in Abhängigkeit ihrer Methylierung

	Expressionslevel							
	exprimiert		in exprimierten Bereich inseriert		stark exprimiert		stark exprimiert kein ORF	
Anzahl von Retrokopien	10		301		119		48	
Methylierung der CGIs	(#)	(%)	(#)	(%)	(#)	(%)	(#)	(%)
CGIs m < 20%	2	20	136	45	57	48	21	44
CGIs 20% ≥ m ≤ 80%	1	10	23	8	12	10	15	31
CGIs m > 80%	2	20	122	41	45	38	11	23
keine Abdeckung	5	50	20	7	5	4	1	2

	Expressionslevel							
	schwach exprimiert		schwach exprimiert kein ORF		Pseudogen		gesamte Anzahl	
Anzahl von Retrokopien	109		50		22		659	
Methylierung der CGIs	(#)	(%)	(#)	(%)	(#)	(%)	(#)	(%)
CGIs m < 20%	14	13	30	60	5	23	265	40
CGIs 20% ≥ m ≤ 80%	12	11	3	6	2	9	68	10
CGIs m > 80%	74	68	10	20	14	64	278	42
keine Abdeckung	9	8	7	14	1	5	48	7

Die Tabelle zeigt den Methylierungsgrad aller *intragenic* R-CGIs in Abhängigkeit der Expression der Retrokopie (UCSC Browser).

A.4.6 Methylierungsgrad in anderen Gewebetypen der 17 Retrokopie-assoziierten CGIs

Tabelle A.20: Methylierungsgrad von zwei Retrokopie-assoziierten CGIs in anderen Gewebetypen

Probenname	Zellen/Gewebe	2391_1_hg19		6141_1_hg19	
		Meth.(%)	Abd.	Meth.(%)	Abd.
Monozyten Methylom 1 ^a	Monozyten	56	12	49	18
Monozyten Methylom 2 ^a	Monozyten	63	8	54	7
Frontal_cortex_normal_1 ^b	Kortex	57	42	32	46
Frontal_cortex_normal_2 ^b	Kortex	47	32	25	35
Frontal_cortex_AD_1 ^b	Kortex	29	41	29	43
Frontal_cortex_AD_2 ^b	Kortex	56	57	35	54
IMR90 ^b	Lunge, fetal, Fibroblasten	75	18	24	12
Colon_Primary_Normal ^b	Darm	70	54	45	33
Human sperm	Sperma ^b	2	6	5	9
Adult liver replicate 1 ^b	Leber	87	63	49	65
Adult liver replicate 2 ^b	Leber	71	49	28	43
Hippocampus middle replicate 1 ^b	Hippokampus (Zentrum)	60	52	35	40
Hippocampus middle replicate 2 ^b	Hippokampus (Zentrum)	52	39	33	35
Fetal heart (119) ^b	fetales Herz	50	40	19	34
Fetal thymus (1238) ^b	fetaler Thymus	45	41	15	36
Fetal adrenal (1244) ^b	fetale Nebenniere	26	25	21	24
Fetal muscle leg (1243) ^b	fetaler Muskel (Bein)	49	35	11	34
Fetal brain (515) ^b	fetales Gehirn	58	24	19	24
Oocytes1 ^c	Oozyten	1	6	5	7
Oocytes2 ^c	Oozyten	3	3	11	3

Die Tabelle zeigt den Methylierungsgrad der R-CGIs 2391_1_hg19 und 6141_1_hg19 in unterschiedlichen Gewebetypen. Angegeben ist jeweils neben dem durchschnittlichen Methylierungslevel (Meth.) in Prozent die durchschnittliche Abdeckung (Abd.) für das jeweilige CGI.

^a veröffentlicht in Rademacher u. a. (2014)

^b veröffentlicht in Ziller u. a. (2013)

^c veröffentlicht in Okae u. a. (2014)

Tabelle A.21: Methylierungsgrad von zehn Retrokopie-assoziierten CGIs in anderen Gewebetypen

Probenname	Zellen/Gewebe	9224_1_hg19 Meth.(%)	9224_1_hg19 Abd.	9261_1_hg19 Meth.(%)	9261_1_hg19 Abd.	9377_1_hg19 Meth.(%)	9377_1_hg19 Abd.	9473_1_hg19 Meth.(%)	9473_1_hg19 Abd.	13085_1_hg19 Meth.(%)	13085_1_hg19 Abd.
Monozyten Methyloom 1 ^a	Monozyten	58	13	72	15	39	7	79	12	76	8
Monozyten Methyloom 2 ^a	Monozyten	31	12	49	6	65	5	80	5	78	3
Frontal_cortex_normal_1 ^b	Kortex	57	51	41	42	63	49	92	45	83	47
Frontal_cortex_normal_2 ^b	Kortex	61	32	45	28	66	25	83	32	90	36
Frontal_cortex_AD_1 ^b	Kortex	69	43	31	38	67	37	82	41	91	35
Frontal_cortex_AD_2 ^b	Kortex	62	67	46	46	67	47	87	55	91	42
IMR90 ^b	Lunge, fetal, Fibroblasten	24	12	56	24	44	15	70	14	91	21
Colon_Primary_Normal ^b	Darm	45	33	55	66	37	39	79	44	84	38
Human sperm	Sperma ^b	0	5	6	6	1	8	2	6	2	9
Adult liver replicate 1 ^b	Leber	66	46	62	57	29	50	80	73	82	68
Adult liver replicate 2 ^b	Leber	79	56	48	69	30	44	68	50	72	53
Hippocampus middle replicate 1 ^b	Hippokampus (Zentrum)	62	41	46	67	60	37	90	60	75	52
Hippocampus middle replicate 2 ^b	Hippokampus (Zentrum)	61	47	42	45	57	38	89	46	72	35
Fetal heart (119) ^b	fetales Herz	66	40	53	69	31	32	66	40	51	35
Fetal thymus (1238) ^b	fetaler Thymus	50	29	44	57	29	40	73	35	77	37
Fetal adrenal (1244) ^b	fetale Nebenniere	76	22	28	29	17	34	65	32	60	29
Fetal muscle leg (1243) ^b	fetaler Muskel (Bein)	61	25	37	40	26	35	77	35	64	41
Fetal brain (515) ^b	fetales Gehirn	66	20	39	28	33	18	89	27	70	21
Oocytes1 ^c	Oozyten	91	9	94	9	2	8	7	7	0	2
Oocytes2 ^c	Oozyten	100	2	100	2	2	2	23	2	12	3
Probenname	Zellen/Gewebe	13250_1_hg19 Meth.(%)	13250_1_hg19 Abd.	15400_1_hg19 Meth.(%)	15400_1_hg19 Abd.	15512_1_hg19 Meth.(%)	15512_1_hg19 Abd.	16448_1_hg19 Meth.(%)	16448_1_hg19 Abd.	16458_1_hg19 Meth.(%)	16458_1_hg19 Abd.
Monozyten Methyloom 1 ^a	Monozyten	22	10	69	16	54	18	79	13	31	11
Monozyten Methyloom 2 ^a	Monozyten	1	4	77	11	35	7	90	10	12	5
Frontal_cortex_normal_1 ^b	Kortex	2	60	88	42	86	28	49	45	18	40
Frontal_cortex_normal_2 ^b	Kortex	4	43	91	29	89	16	52	44	11	33
Frontal_cortex_AD_1 ^b	Kortex	6	47	85	37	70	20	68	38	40	31
Frontal_cortex_AD_2 ^b	Kortex	12	68	80	45	83	27	56	53	40	38
IMR90 ^b	Lunge, fetal, Fibroblasten	8	16	70	22	80	21	54	16	15	14
Colon_Primary_Normal ^b	Darm	7	44	86	52	74	51	88	61	43	38
Human sperm	Sperma ^b	0	9	3	6	8	5	8	11	2	8
Adult liver replicate 1 ^b	Leber	10	57	69	69	88	68	72	76	31	43
Adult liver replicate 2 ^b	Leber	5	46	70	49	73	56	47	64	37	47
Hippocampus middle replicate 1 ^b	Hippokampus (Zentrum)	10	48	89	66	89	60	56	59	14	39
Hippocampus middle replicate 2 ^b	Hippokampus (Zentrum)	2	45	92	44	83	36	50	44	36	34
Fetal heart (119) ^b	fetales Herz	5	37	79	48	38	36	31	44	15	31
Fetal thymus (1238) ^b	fetaler Thymus	2	37	68	35	73	33	90	34	25	24
Fetal adrenal (1244) ^b	fetale Nebenniere	2	30	69	27	69	24	28	22	20	21
Fetal muscle leg (1243) ^b	fetaler Muskel (Bein)	1	38	86	37	59	26	37	31	22	24
Fetal brain (515) ^b	fetales Gehirn	0	25	90	27	55	21	40	27	20	24
Oocytes1 ^c	Oozyten	0	4	0	10	94	6	2	5	0	8
Oocytes2 ^c	Oozyten	1	3	0	2	100	1	2	2	9	3

Die Tabelle zeigt den Methylierungsgrad von zehn R-CGIs (9224_1_hg19, 9261_1_hg19, 9377_1_hg19, 9473_1_hg19, 13085_1_hg19, 13250_1_hg19, 15400_1_hg19, 15512_1_hg19, 16448_1_hg19 und 16458_1_hg19) in unterschiedlichen Gewebetypen. Angegeben ist jeweils neben dem durchschnittlichen Methylierungslevel (Meth.) in Prozent die durchschnittliche Abdeckung (Abd.) für das jeweilige CGI.

^a veröffentlicht in Rademacher u. a. (2014)

^b veröffentlicht in Ziller u. a. (2013)

^c veröffentlicht in Okae u. a. (2014)

Tabelle A.22: Methylierungsgrad von fünf Retrogen-überlappenden CGIs in anderen Geweben

Probenname	Zellen/Gewebe	17031_1_hg19 Meth.(%)	17031_1_hg19 Abd.	20403_1_hg19 Meth.(%)	20403_1_hg19 Abd.	23548_1_hg19 Meth.(%)	23548_1_hg19 Abd.	24982_1_hg19 Meth.(%)	24982_1_hg19 Abd.	26859_1_hg19 Meth.(%)	26859_1_hg19 Abd.
Monozyten	Monozyten	42	7	20	13	59	14	55	20	77	8
Monozyten	Monozyten	46	18	13	11	42	10	55	12	94	4
Frontal_cortex_normal_1 ^b	Kortex	65	40	3	32	81	51	27	44	73	61
Frontal_cortex_normal_2 ^b	Kortex	57	31	1	24	78	36	34	35	73	50
Frontal_cortex_AD_1 ^b	Kortex	35	30	4	30	68	38	29	38	71	41
Frontal_cortex_AD_2 ^b	Kortex	52	51	3	41	58	47	35	59	73	72
IMR90 ^b	Lunge, fetal, Fibroblasten	33	14	12	9	48	16	57	26	61	18
Colon_Primary_Normal ^b	Darm	72	46	22	30	77	65	55	63	86	26
Human sperm	Sperma ^b	8	10	3	8	3	9	5	6	5	3
Adult liver replicate 1 ^b	Leber	64	56	24	30	44	79	58	94	83	41
Adult liver replicate 2 ^b	Leber	55	56	16	32	41	51	59	62	79	33
Hippocampus middle replicate 1 ^b	Hippokampus (Zentrum)	58	41	5	38	59	79	43	83	88	41
Hippocampus middle replicate 2 ^b	Hippokampus (Zentrum)	54	38	7	30	63	58	39	64	88	27
Fetal heart (119) ^b	fetales Herz	41	38	5	30	29	38	51	53	78	24
Fetal thymus (1238) ^b	fetaler Thymus	36	31	9	29	40	31	44	47	67	38
Fetal adrenal (1244) ^b	fetale Nebenniere	27	25	5	22	23	33	46	36	85	12
Fetal muscle leg (1243) ^b	fetaler Muskel (Bein)	16	36	6	25	56	38	45	44	68	38
Fetal brain (515) ^b	fetales Gehirn	45	26	4	16	31	30	55	34	71	25
Oocytes1 ^c	Oozyten	0	4	7	6	17	6	83	3	4	9
Oocytes2 ^c	Oozyten	13	1	0	3	2	2	89	5	3	4

Die Tabelle zeigt die Methylierungslevel von fünf RO-CGIs (17031_1_hg19, 20403_1_hg19, 23548_1_hg19, 24982_1_hg19 und 26859_1_hg19) in unterschiedlichen Gewebetypen. Angegeben ist jeweils neben dem durchschnittlichen Methylierungslevel (Meth.) in Prozent die durchschnittliche Abdeckung (Abd.) für das jeweilige CGI.

^a veröffentlicht in Rademacher u. a. (2014)

^b veröffentlicht in Ziller u. a. (2013)

^c veröffentlicht in Okae u. a. (2014)

A.4.7 Methylierungsmuster der 12 R-CGIs mittels Roche/454 GS Junior

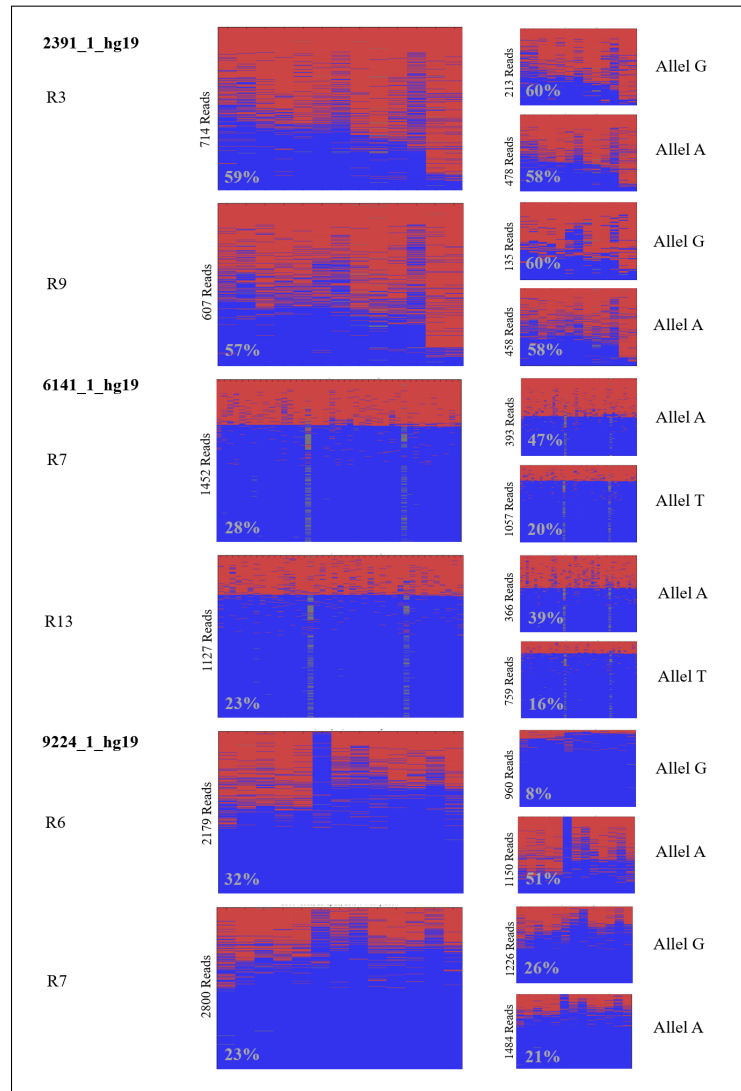


Abbildung A.13: Methylierungsanalyse einzelner CpGs der 12 R-CGIs in Monozyten von heterozygoten Individuen - Part 1

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.2.2.4 Tabelle 3.8. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 1 zeigt die Methylierungsmuster der CGIs 2391_1_hg19, 6141_1_hg19 und 9224_1_hg19.

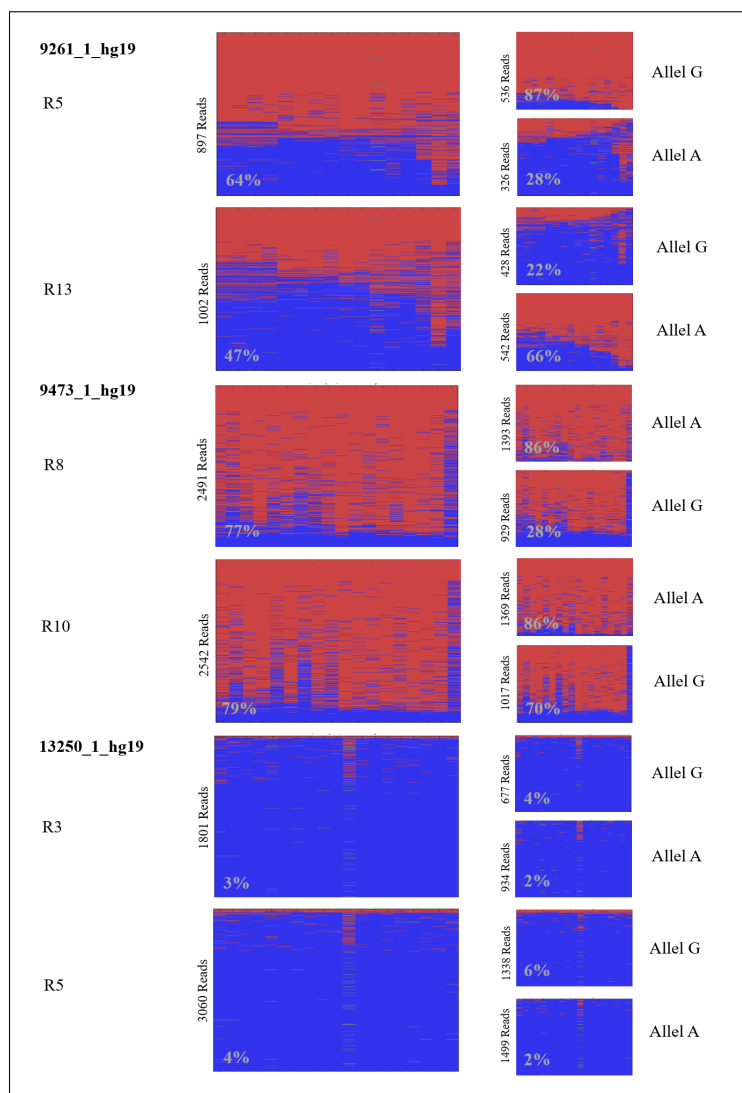


Abbildung A.14: Methylierungsanalyse einzelner CpGs der 12 R-CGIs in Monozyten von heterozygoten Individuen - Part 2

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.2.2.4 Tabelle 3.8. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 2 zeigt die Methylierungsmuster der CGIs 9261_1_hg19, 9473_1_hg19 und 13250_1_hg19.

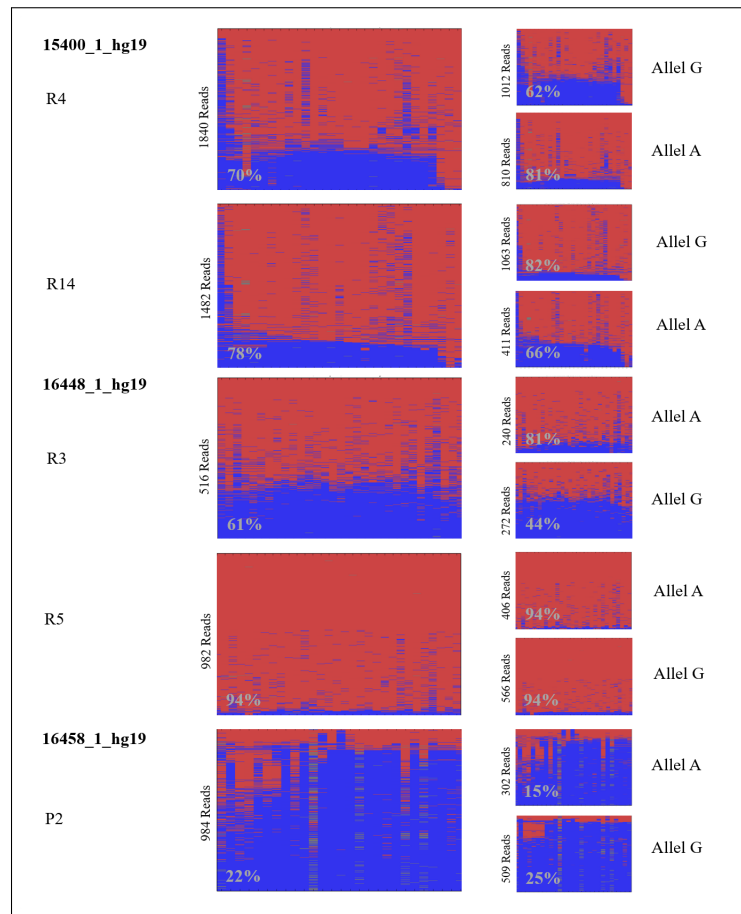


Abbildung A.15: Methylierungsanalyse einzelner CpGs der 12 R-CGIs in Monozyten von heterozygoten Individuen - Part 3

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.2.2.4 Tabelle 3.8. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 3 zeigt die Methylierungsmuster der CGIs 15400_1_hg19, 16448_1_hg19 und 16458_1_hg19.

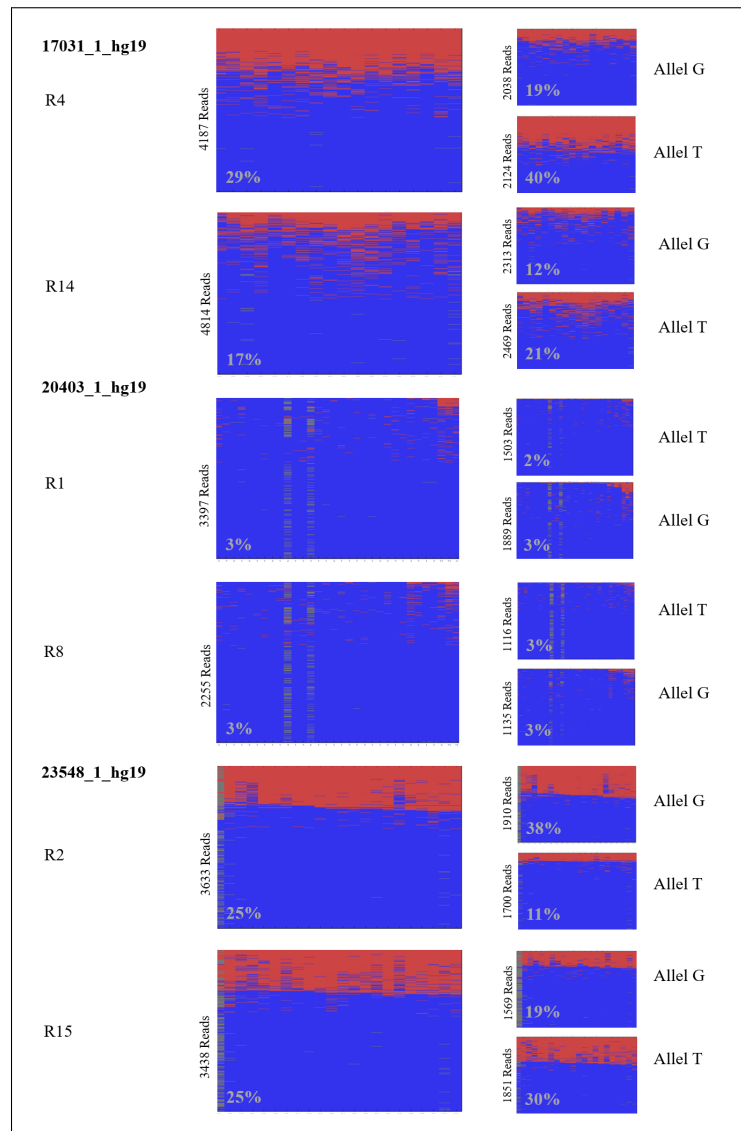


Abbildung A.16: Methylierungsanalyse einzelner CpGs der 12 R-CGIs in Monozyten von heterozygoten Individuen - Part 4

Die Abbildung zeigt die Methylierungsplots aller untersuchten Proben von Abschnitt 3.2.2.4 Tabelle 3.8. Die Methylierungsplots auf der linken Seite zeigen die Methylierung aller Reads eines Amplikons und die Plots auf der rechten Seite die Methylierung aufgetrennt nach Allelen. In jedem Plot ist die durchschnittliche Methylierung aller Reads über alle CpGs angegeben (grau). Methylierte CpGs sind in rot dargestellt und unmethylierte CpGs in blau. Part 4 zeigt die Methylierungsmuster der CGIs 17031_1_hg19, 20403_1_hg19 und 23548_1_hg19.

A.5 Evolutionsanalyse

Tabelle A.23: Evolutionsanalyse von 104 human/nicht-murinen intronischen CGIs mit zusätzlichen Treffern im Genom. (*gap*) Sequenzlücke im Genom, * Teile des Gens nicht vorhanden. Diese Tabelle wurde schon in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

CGI ID	hg19	panTro4	gorGor3	ponAbe2	nomLeu3	rheMac3	calJac3	otoGar3	mm10
334_1_hg19	ja	ja	ja	ja	ja	ja	ja	ja	nein
1094_1_hg19	ja	ja	ja	ja	ja	nein	nein	nein	nein
1147_1_hg19	ja	ja	ja	ja	ja	ja	ja	ja	nein
1227_1_hg19	ja	ja	nein*	ja	ja	ja	nein(<i>gap</i>)	nein	nein
1255_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
1287_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
1330_1_hg19	ja	ja	ja	ja	ja	ja	ja	ja	nein
1362_1_hg19	ja	ja	ja	ja	ja	nein	nein	nein	nein
1720_1_hg19	ja	ja	ja	ja	ja	ja	ja	ja	nein
1909_1_hg19	ja	ja	ja	nein	nein	nein	nein	nein(<i>gap</i>)	nein
1911_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
2392_1_hg19	ja	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein	nein
2708_1_hg19	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein	nein(<i>gap</i>)	nein
3524_1_hg19	ja	ja	ja	nein(<i>gap</i>)	ja	nein*	nein	nein	nein
3783_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein*	nein
3926_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
4213_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
4675_1_hg19	ja	ja	ja	ja	ja	nein	nein(<i>gap</i>)	nein	nein
4754_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
5128_1_hg19	ja	ja	ja	nein	ja	ja	ja	nein	nein
5148_1_hg19	ja	ja	ja	nein*	ja	nein(<i>gap</i>)	nein(<i>gap</i>)	nein	nein
5707_1_hg19	ja	ja	nein	nein	nein	nein	nein	nein	nein
5734_1_hg19	ja	ja	ja	ja	ja	nein	nein	nein	nein
5735_1_hg19	ja	ja	nein(<i>gap</i>)	ja	ja	nein	nein	nein	nein
5774_1_hg19	ja	ja	ja	ja	ja	nein	nein	nein	nein
5895_1_hg19	ja	ja	ja	nein	nein	nein	nein	nein	nein
6119_1_hg19	ja	ja	nein	nein	nein	nein	nein	nein	nein
6360_1_hg19	ja	ja	ja	ja	ja	ja*	ja	nein	nein
6681_1_hg19	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein	nein	nein
6781_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
6782_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
6931_1_hg19	ja	ja	nein	nein	nein	nein	nein	nein	nein
6952_1_hg19	ja	ja	ja	ja	ja	nein	nein*	nein	nein
7353_1_hg19	ja	ja	nein*	ja	ja	ja	nein*	nein(<i>gap</i>)	nein
7430_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
7480_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
8436_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
9007_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
9008_1_hg19	ja	ja	ja	ja	ja	ja	ja	ja	nein
9009_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
9933_1_hg19	ja	ja	ja	ja	ja	nein*	nein	nein	nein
9934_1_hg19	ja	ja	ja	ja	ja	nein*	nein(<i>gap</i>)	nein	nein
10233_1_hg19	ja	nein*	nein*	ja	nein*	nein*	nein	nein*	nein
10850_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
11376_1_hg19	ja	ja	ja	ja	nein(<i>gap</i>)	nein(<i>gap</i>)	ja	ja	nein
11477_1_hg19	ja	ja	ja	ja	ja	nein(<i>gap</i>)	ja	nein	nein
11479_1_hg19	ja	ja	ja	ja	ja	ja	nein*	nein*	nein
11481_1_hg19	ja	ja	ja	ja	ja	ja	nein*	nein*	nein
11497_1_hg19	ja	ja	ja	nein	nein	nein	nein	nein	nein
11511_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
11525_1_hg19	ja	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein(<i>gap</i>)	nein
12025_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
13648_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein
14414_1_hg19	ja	ja	nein(<i>gap</i>)	ja	ja	nein*	nein*	nein*	nein
14676_1_hg19	ja	ja	ja	ja	ja	nein*	ja	nein*	nein
14830_1_hg19	ja	ja	ja	ja	ja	nein	nein	nein	nein
14831_1_hg19	ja	ja	ja	ja	ja	nein	nein	nein	nein
15161_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
15180_1_hg19	ja	ja	nein*	ja	ja	nein*	ja	nein	nein
15205_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
15224_1_hg19	ja	ja	ja	ja	ja	nein	nein	nein	nein
15290_1_hg19	ja	ja	ja	nein	nein	nein	nein	nein	nein
15352_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
15356_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
15798_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
16205_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
16226_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein
16253_1_hg19	ja	ja	nein(<i>gap</i>)	ja	ja	ja	ja	nein	nein
16256_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
16633_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
16634_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
16694_1_hg19	ja	ja	ja	nein	nein	nein	nein	nein	nein
17304_1_hg19	ja	ja	ja	nein	nein	nein	nein	nein	nein
17484_1_hg19	ja	ja	ja	nein	nein	nein	nein	nein	nein
17523_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
17547_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
18132_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein

Tabelle A.23: Evolutionsanalyse von 104 human/nicht-murinen intronischen CGIs mit zusätzlichen Treffern im Genom. (*gap*) Sequenzlücke im Genom, * Teile des Gens nicht vorhanden. Diese Tabelle wurde schon in abgeänderter Form in Rademacher u. a. (2014) veröffentlicht.

CGI ID	hg19	panTro4	gorGor3	ponAbe2	nomLeu3	rheMac3	calJac3	otoGar3	mm10
18136_1_hg19	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein	nein	nein
18308_1_hg19	ja	nein(<i>gap</i>)	nein(<i>gap</i>)	nein	nein	nein	nein	nein	nein
18512_1_hg19	ja	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein	nein
19100_1_hg19	ja	nein*	nein(<i>gap</i>)	nein*	nein	nein	nein	nein	nein
19854_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
19870_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein*	nein
19877_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
19878_1_hg19	ja	ja	nein	nein	nein	nein	nein	nein	nein
19881_1_hg19	ja	ja	ja	ja	ja	nein	nein	nein	nein
19883_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
20085_1_hg19	ja	ja	ja	nein*	ja	ja	nein(<i>gap</i>)	nein(<i>gap</i>)	nein
20632_1_hg19	ja	ja	nein*	nein	nein	nein*	nein	nein	nein
20636_1_hg19	ja	ja	nein	nein	nein	nein	nein	nein	nein
20657_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein(<i>gap</i>)	nein
21034_1_hg19	ja	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein*	nein
21352_1_hg19	ja	ja	ja	nein	nein	nein*	nein*	nein	nein
21920_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein
22285_1_hg19	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein(<i>gap</i>)	nein	nein
25767_1_hg19	ja	nein	nein	nein	nein	nein	nein	nein	nein
26094_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
26614_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein	nein
26734_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein*	nein
26772_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
26785_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein(<i>gap</i>)	nein
26912_1_hg19	ja	ja	ja	ja	ja	ja	nein	nein	nein
26939_1_hg19	ja	nein(<i>gap</i>)	ja	ja	ja	ja	ja	ja	nein
27070_1_hg19	ja	ja	ja	ja	ja	ja	ja	nein*	nein

A.6 RePCR-Programme

Bei den CGIs 9261_1_hg19 und 15400_1_hg19 wurde anstatt der *twostep* RePCR (2.4.2.3) das folgende PCR-Programm als RePCR angewendet:

15min	95°C	(Polymeraseaktivierung)	
30sek	95°C	(Denaturierung)	} 40 Zyklen
30sek	59°C	(Annaeling)	
45sek	72°C	(Elongation)	
10min	72°C	(finale Elongation)	
∞	4°C		

A.7 Ansätze und PCR-Programme der RT-PCR

Tabelle A.24: RT-PCR Ansätze von *RPS2P32*, *IGF2BP3* und *β-Aktin*

		<i>RPS2P32</i>	<i>IGF2BP3</i>	<i>β-Aktin</i>
dNTPs	1,25mM	8µl	8µl	8µl
GoTaq Green Puffer	5x	10µl	10µl	
Puffer	10x			5µl
MgCl ₂	25mM	3µl	3µl	3µl
For-Primer	20µM	1µl	1µl	1µl
Rev-Primer	20µM	1µl	1µl	1µl
GoTaq	5U/µl	0,5µl	1µl	
Gold-Taq	5U/µl			0,5µl
H ₂ O-Merck		20,5µl	10µl	25,5µl
RT-Ansatz		6µl	16µl	6µl
Summe		50µl	50µl	50µl

PCR-Programm *RPS2P32*:

2min	95°C	(Polymeraseaktivierung)	
30sek	95°C	(Denaturierung)	} 40 Zyklen
30sek	68°C	(Annaeling)	
45sek	72°C	(Elongation)	
5min	72°C	(finale Elongation)	
∞	4°C		

PCR-Programm *IGF2BP3*:

2min	95°C		
30sek	95°C	}	14 Zyklen
30sek	68°C		
30sek	72°C		
30sek	95°C	}	45 Zyklen
30sek	62°C		
30sek	72°C		
5min	72°C		
∞	4°C		

PCR-Programm *β-Aktin*:

10min	95°C	(Polymeraseaktivierung)	
20sek	95°C	(Denaturierung)	}
20sek	56°C	(Annaeling)	
45sek	72°C	(Elongation)	
10min	72°C	(finale Elongation)	
∞	4°C		

A.8 Zusätzliche Plots zu den Methylierungsanalysen der Umgebung von Retrokopie-assoziierten CGIs unter Berücksichtigung von repetitiven Sequenzen

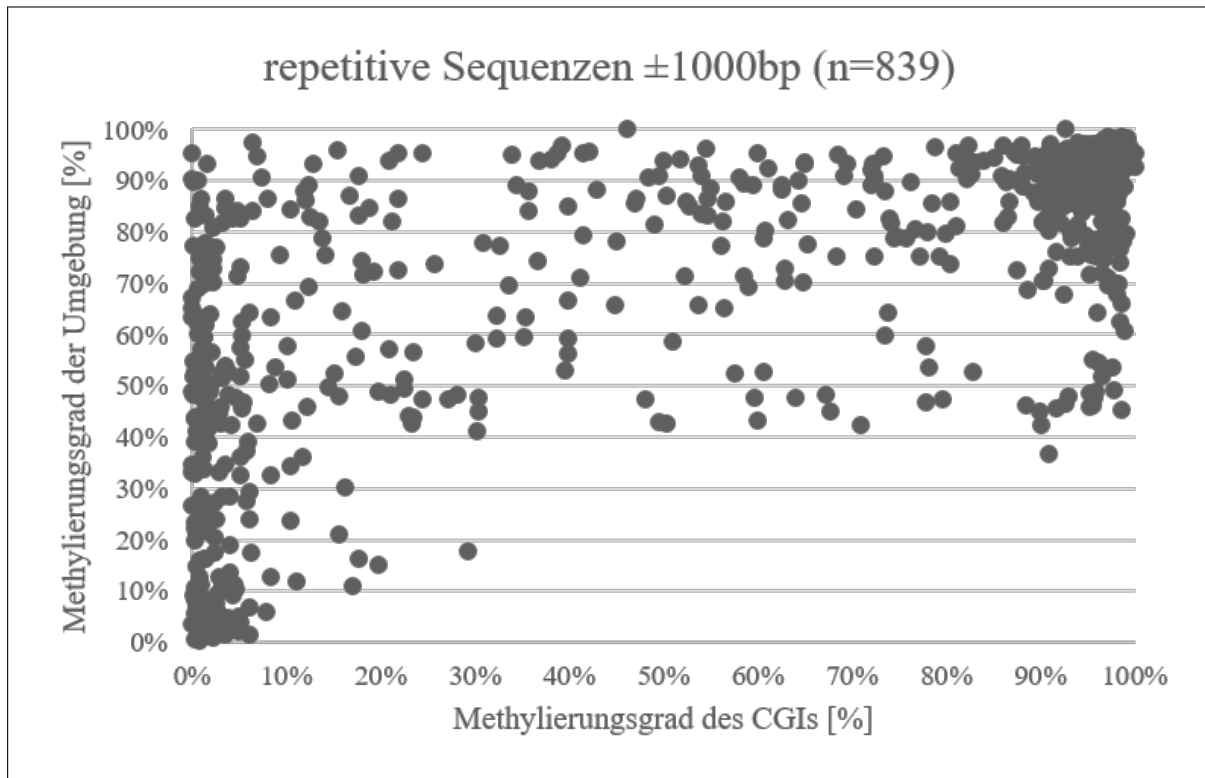


Abbildung A.17: Methylierungsplot vom Methylierungsgrad der R-CGIs in Abhängigkeit vom Methylierungsgrad der Umgebung der Retrokopie, in der sowohl 1000 bp *upstream* als auch *downstream* repetitive Sequenzen liegen

Die Abbildung zeigt den Grad der Methylierung der R-CGIs in Abhängigkeit des Methylierungsgrads der Umgebung der Retrokopie. Jeder einzelne Punkt repräsentiert den Methylierungsgrad eines R-CGIs (X-Achse) und den dazugehörigen Methylierungsgrad der Umgebung der Retrokopie (Y-Achse) in Prozent, wobei sowohl 1000 bp *upstream* als auch *downstream* der Retrokopie repetitive Sequenzen liegen. Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

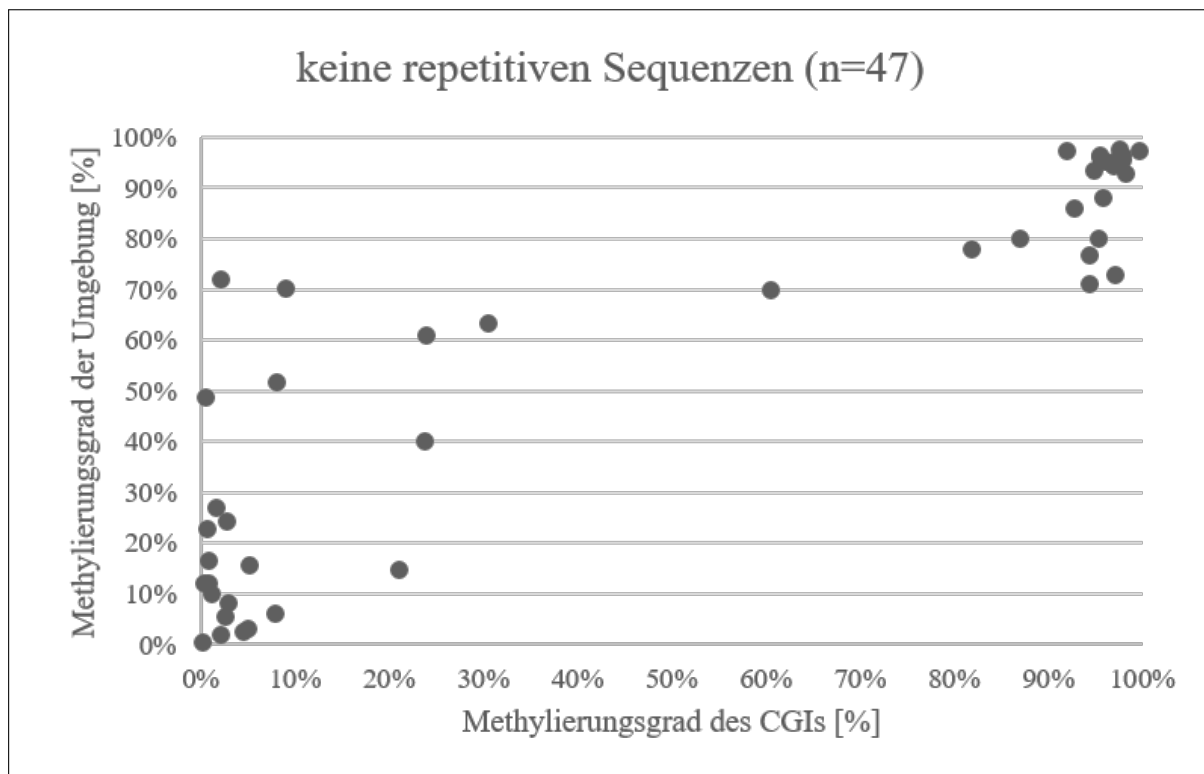


Abbildung A.18: Methylierungsplot vom Methylierungsgrad der R-CGIs in Abhängigkeit vom Methylierungsgrad der Umgebung der Retrokopie, in der keine repetitive Sequenzen liegen

Die Abbildung zeigt den Grad der Methylierung der R-CGIs in Abhängigkeit des Methylierungsgrads der Umgebung der Retrokopie. Jeder einzelne Punkt repräsentiert den Methylierungsgrad eines R-CGIs (X-Achse) und den dazugehörigen Methylierungsgrad der Umgebung der Retrokopie (Y-Achse) in Prozent, wobei in der Umgebung der Retrokopie keine repetitive Sequenzen liegen. Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

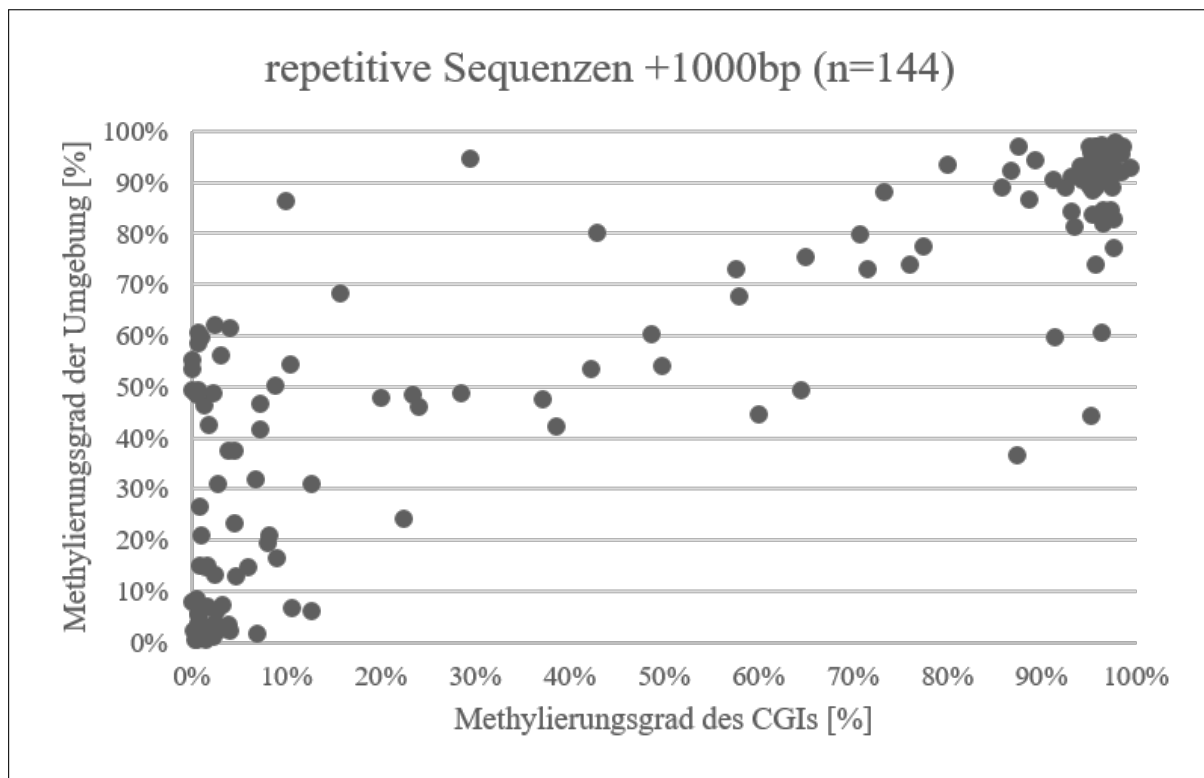


Abbildung A.19: Methylierungsplot vom Methylierungsgrad der R-CGIs in Abhängigkeit vom Methylierungsgrad der Umgebung der Retrokopie, wobei 1000 bp *upstream* der Retrokopie repetitive Sequenzen liegen

Die Abbildung zeigt den Grad der Methylierung der R-CGIs in Abhängigkeit des Methylierungsgrads der Umgebung der Retrokopie. Jeder einzelne Punkt repräsentiert den Methylierungsgrad eines R-CGIs (X-Achse) und den dazugehörigen Methylierungsgrad der Umgebung der Retrokopie (Y-Achse) in Prozent, wobei nur 1000 bp *upstream* der Retrokopie repetitive Sequenzen liegen. Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

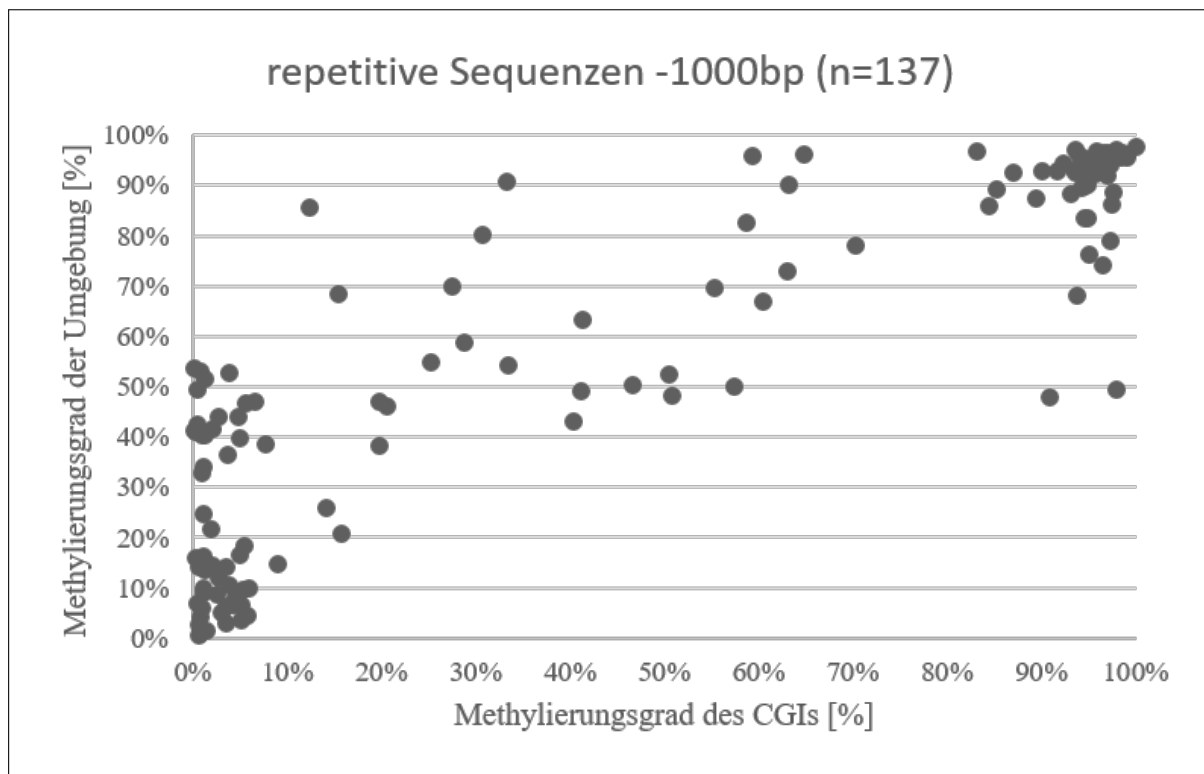


Abbildung A.20: Methylierungsplot vom Methylierungsgrad der R-CGIs in Abhängigkeit vom Methylierungsgrad der Umgebung der Retrokopie, wobei 1000 bp *downstream* der Retrokopie repetitive Sequenzen liegen

Die Abbildung zeigt den Grad der Methylierung der R-CGIs in Abhängigkeit des Methylierungsgrads der Umgebung der Retrokopie. Jeder einzelne Punkt repräsentiert den Methylierungsgrad eines R-CGIs (X-Achse) und den dazugehörigen Methylierungsgrad der Umgebung der Retrokopie (Y-Achse) in Prozent, wobei nur 1000 bp *downstream* der Retrokopie repetitive Sequenzen liegen. Diese Abbildung wurde in abgeänderter Form bereits in Grothaus u. a. (2016) veröffentlicht.

A.9 Rohdaten der SNaPshot-Analyse mit anschließendem Genescan

Tabelle A.25: Rohdaten der quantitativen Analyse des allelischen Transkriptverhältnisses zu *RPS2P32*

Probenname	Peak area	A mat / G pat	Rate 1	G pat / A mat	Rate 2
Kind Trio III DNA G Allel	52038	0,40	1	2,47	1
Kind Trio III DNA A Allel	21063				
Kind Trio III cDNA G Allel F	52038	0,34	0,85	2,91	1,18
Kind Trio III cDNA A Allel F	4816				
Kind Trio III cDNA G Allel R	22161	0,23	0,56	4,39	1,78
Kind Trio III cDNA A Allel R	5048				

Die Tabelle zeigt die Ergebnisse der quantitativen Analyse des allelischen Transkriptverhältnisses von Abschnitt 3.2.3.4. Neben dem Probennamen ist die *Peak area*, das Verhältnis A mat zu G mat mit dazugehöriger Rate und das Verhältnis von G pat zu A mat mit dazugehöriger Rate angegeben. Diese Tabelle wurde schon in abgeänderter Form in Grothaus u. a. (2016) veröffentlicht.

Tabelle A.26: Rohdaten der quantitativen Analyse des allelischen Transkriptverhältnisses des *IGF2BP3*-Gens

Probenname	Peak area	A pat / G mat	Rate 1	G mat / A pat	Rate 2
Kind A DNA G Allel	39096	0,55	1	1,83	1
Kind A DNA A Allel	21421				
Kind A cDNA G Allel	18561	0,52	0,95	1,91	1,05
Kind A cDNA A Allel	9721				
Kind B DNA G Allel	20233	0,54	1	1,84	1
Kind B DNA A Allel	11004				
Kind B cDNA G Allel	11293	0,65	1,20	1,54	0,84
Kind B cDNA A Allel	7322				

Die Tabelle zeigt die Ergebnisse der quantitativen Analyse des allelischen Transkriptverhältnisses des *IGF2BP3*-Gens von Abschnitt 3.2.3.5 Abbildung 3.16. Neben dem Probennamen ist die *Peak area*, das Verhältnis A mat zu G mat mit dazugehöriger Rate und das Verhältnis von G pat zu A mat mit dazugehöriger Rate angegeben.

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die zum Gelingen dieser Arbeit beigetragen haben. Mein herzlicher Dank gilt vor allem:

Herrn **Prof. Dr. Bernhard Horsthemke** für die Möglichkeit meine Dissertation am Institut für Humangenetik anzufertigen, für das interessante Forschungsprojekt, für die hervorragende wissenschaftliche Betreuung sowie für die wertvollen Ideen und Anregungen.

Frau **Dr. Deniz Kanber** für die freundschaftliche Betreuung, für die geduldige Einarbeitung in die Weiten der Laborwelt, für die Teilung des Laborplatzes, für die Beantwortung von zahlreichen Fragen und vor allem für die zahlreichen Stunden, die sie in mein (unser) Projekt gesteckt hat.

Allen Mitarbeitern des Instituts (jetzigen und ehemaligen), vor allem **Barbara, Christian, Christina, Dagmar, Dani, Hannah, Helena, Jana, Jasmin, Karin, Lars, Laura, Melanie, Michael** und **Sabine** für die schöne Arbeitsatmosphäre, für die gute Zusammenarbeit, für die Unterstützung und für viele gute Ratschläge und Ideen. Ganz besonders danke ich meinen beiden Büromitbewohnern **Stephanie** und **Hermann-Josef** für viele spannende Diskussionen, für die freundschaftliche Unterstützung, für die sehr gute Verpflegung im Büro und für die stets offenen Ohren.

Mein besonderer Dank gilt all den Menschen, die mich auf dem Weg hierher begleitet haben, **meiner Familie** und **meinen Freunden**, für ihre Unterstützung, für ihren Zuspruch und für willkommene Ablenkungen vom Arbeitsalltag.

Nils gebührt mein größtmöglicher Dank für seine Unterstützung in allen Lebenslagen, für seine Geduld, für sein Vertrauen in mich und für einfach alles.

B Über die Autorin

B.1 Lebenslauf

„Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.“

B.2 Publikationen

Grothaus K, Kanber D, Gellhaus A, Mikat B, Kolarova J, Siebert R, Wieczorek D, Horsthemke B. Genome-wide methylation analysis of retrocopy associated CpG islands and their genomic environment. *Epigenetics*. 2016. doi: 10.1080/15592294.2016.1145330

Grosser C, Wagner N, **Grothaus K**, Horsthemke B. Altering TET dioxygenase levels within physiological range affects DNA methylation dynamics of HEK293 cells. *Epigenetics* 2015. doi: 10.1080/15592294.2015.1073879.

Berulava T, Rahmann S, **Rademacher K**, Klein-Hitpass L, Horsthemke B. N6-Adenosine Methylation in MiRNAs. *PLoS ONE* 2015, 10(2): e0118438. doi: 10.1371/journal.pone.0118438.

Dankert D, Demond H, Tropphoff T, Heiligentag M, **Rademacher K**, Eichenlaub-Ritter U, Horsthemke B, Grümmer R. Pre- and postovulatory aging of murine oocytes affect the transcript level and poly(A) tail length of maternal effect genes. *PLoS ONE* 2014, 9(10):e108907. doi: 10.1371/journal.pone.0108907.

Rademacher K, Schröder C, Kanber D, Klein-Hitpass L, Wallner S, Zeschnigk M, Horsthemke B. Evolutionary origin and methylation status of human intronic CpG islands that are not present in mouse. *Genome Biol. Evol.* 2014, 6(7):1579-88. doi.1093/gbe/evu125

Beygo J, Ammerpohl O, Gritzan D, Heitmann M, **Rademacher K**, Richter J, Caliebe A, Siebert R, Horsthemke B, Buiting K. Deep bisulfite sequencing of aberrantly methylated Loci in a patient with multiple methylation defects. *PLoS ONE* 2013, 8(10):e76953.

Beygo J, Citro V, Sparago A, De Crescenzo A, Cerrato F, Heitmann M, **Rademacher K**, Guala A, Enklaar T, Anichini C, et al. The molecular function and clinical phenotype of partial deletions of the IGF2/H19 imprinting control region depends on the spatial arrangement of the remaining CTCF-binding sites. *Hum Mol Genet* 2013; 22:544-57.

Kosubek A, Klein-Hitpass L, **Rademacher K**, Horsthemke B, Ryffel GU. Aging of *Xenopus tropicalis* eggs leads to deadenylation of a specific set of maternal mRNAs and loss of developmental potential. *PLOS ONE* 2010, 5(10):e13532.

Baumbach J, Wittkop T, **Rademacher K**, Rahmann S, Brinkrolfs K, Tauch A. Coryne-RegNet 3.0—an interactive systems biology platform for the analysis of gene regulatory networks in corynebacteria and Escherichia coli. *J Biotechnol* 2007, 129(2):279-89.

C Erklärungen

Erklärung

Hiermit erkläre ich, gem. §6 Abs. 2, g der Promotionsordnung der Fakultät für Biologie zur Erlangung der Dr. rer. nat., dass ich das Arbeitsgebiet, dem das Thema „Genomweite Untersuchungen zur Evolution und Methylierung von CpG-Inseln“ zuzuordnen ist, in Forschung und Lehre vertrete und den Antrag von Katrin Grothaus befürworte.

Essen, den

Unterschrift d. wissenschaftl. Betreuers/
Mitglieds der Universität Duisburg-Essen
Prof. Dr. B. Horsthemke

Erklärung

Hiermit erkläre ich, gem. §7 Abs. 2, d und f der Promotionsordnung der Fakultät für Biologie zur Erlangung des Dr. rer. nat., dass ich die vorliegende Dissertation selbständig verfasst und mich keiner anderen als der angegebenen Hilfsmittel bedient, bei der Abfassung der Dissertation nur die angegebenen Hilfsmittel benutzt und alle wörtlich oder inhaltlich übernommenen Stellen als solche gekennzeichnet habe.

Essen, den

Unterschrift der Doktorandin (Katrin Grothaus)

Erklärung

Hiermit erkläre ich, gem. §7 Abs. 2, e und g der Promotionsordnung der Fakultät für Biologie zur Erlangung der Dr. rer. nat., dass ich keine anderen Promotionen bzw. Promotionsversuche in der Vergangenheit durchgeführt habe und dass diese Arbeit von keiner anderen Fakultät/Fachbereich abgelehnt worden ist.

Essen, den

Unterschrift der Doktorandin (Katrin Grothaus)